PROGRESS IN BIOPHYSICS

AND BIOPHYSICAL CHEMISTRY

1

Editors

J. A. V. BUTLER

Chester Beatty Research Institute
Royal Cancer Hospital
London

J. T. RANDALL, F. R. S.

Wheatstone Professor of Physics
in the University of London
at King's College

PERGAMON PRESS

OXFORD · LONDON · NEW YORK · PARIS

PERGAMON PRESS LTD.
Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1

PERGAMON PRESS INC.
122 East 55th Street, New York 22, N.Y.
1404 New York Avenue N.W., Washington 5 D.C.

PERGAMON PRESS S.A.R.L.
*24 Rue des Écoles, Paris V^e^*

PERGAMON PRESS G.m.b.H.
Kaiserstrasse 75, Frankfurt am Main

First Published 1950
Second Impression 1961

Originally set and printed in Great Britain
Reprinted by photolithography in Hungary, 1961.

CONTENTS

PAGE

PREFACE *vii*

ACKNOWLEDGMENTS *viii*

1 PROPERTIES OF SOLUTIONS OF LARGE MOLECULES 1
H. GUTFREUND, Department of Colloid Science, The University, Cambridge

2 FUNDAMENTAL STRUCTURES IN BIOLOGICAL SYSTEMS 39
K. M. RUDALL; Textile Physics Research Laboratory, The University, Leeds

3 SCATTERING OF VISIBLE LIGHT AND X-RAYS BY SOLUTIONS OF PROTEINS 73
G. OSTER, Davy Faraday Laboratory, Royal Institution, London

4 BIOELECTRIC POTENTIALS, THEIR MAINTENANCE AND FUNCTION 85
E. E. CRANE, Medical Research Council Unit for Research in Cell Metabolism, Sheffield

5 PHASE-CONTRAST MICROSCOPY IN BIOLOGICAL RESEARCH 137
A. F. W. HUGHES, Strangeways Research Laboratory, Cambridge

6 LOCAL REFRACTOMETRY: THE OBSERVATION OF TRANSPARENT INHOMOGENEOUS MATTER 142
J. St. L. PHILPOT, Radiobiological Research Unit, Atomic Energy Research Establishment, Harwell

7 USE OF SOFT X-RAYS IN THE ASSAY OF BIOLOGICAL MATERIALS 164
A. ENGSTRÖM, Medecinska, Nobel Institutet, Carolinska Institutet, Stockholm

8 THE TOLERANCE OF MAN FOR RADIOACTIVE ISOTOPES 197
J. F. LOUTIT, Radiobiological Research Unit, Atomic Energy Research Establishment, Harwell

9 MECHANICAL PROPERTIES OF FIBRES AND MUSCLES 216
M. G. M. PRYOR, Zoology Department, The University, Cambridge

NAME INDEX 269

SUBJECT INDEX 276

NAME INDEX

Abbot, B. C. 265, 268
Abramson, H. A. 103, 131
Adair, G.S. 1, 14, 16, 17, 19, 21, 22, 32, 37, 38, 119, 131, 135
Adair, M. E. 21, 22, 32, 37, 38
Aird, R. B. 133
Alberty, R. A. 12, 22, 37
Alexander, A. E. 14, 37
Alfrey, T. 31, 38, 235, 267
Allan, E. 132
Allison, S. K. 166, 167, 195
Altmann, M. 107, 131
Ambrose, E. J. 72, 163
Amlong, H. U. 128, 131
Anderson, E. A. 12, 37
Andersson, K. J. J. 23, 38
Ashley-Montague, M. F. 108, 131
Astbury, W. T. 50, 58, 71, 72, 223, 227, 228, 239, 247, 252, 267, 268
Aub, J. C. 207, 209, 215
Auger, D. 112, 131
Austin, C. R. 139, 141
Axelrod, D. J. 215

Baez, A. V. 169, 195
Bailey, K. 26, 38, 72
Baker, B. B. 163
Baldwin, E. 120, 131
Ball, E. G. 133
Barclay, A. E. 170, 196
Barcroft, J. 119, 135
Barer, R. 137, 140, 141
Barker, D. E. 136
Barker, H. H. 207, 208, 215
Barnes, T. C. 100, 110, 131
Barron, E. S. G. 131
Barton, D. S. 106, 107, 132
Barth, L. G. 132
Batchelder, A. C. 21, 22, 37
Bates, M. I. 213, 215
Bate Smith, E. C. 263, 268
Behrens, B. 205, 215
Belar, K. 138, 140
Bell, D. J. 26, 38
Bell, F. O. 72, 228, 267
Bendall, J. R. 263, 268
Bennett, A. H. 137, 140
Bergmann, M. 133
Bernal, J. D. 72, 83, 84
Bernfeld, P. 99, 135
Bernstein, J. 98, 132
Bertholon, M. 104, 123
Berti, L. 137, 140
Bethe, A. 113, 132
Beutner, R. 95, 100, 108, 131, 132
Biedermann, W. 87, 132
Bier, M. 84
Bishop, G. H. 106, 132
Black, V. S. 121, 132
Blaker, R. H. 84
Blinks, L. R. 102, 124, 132
Bloomfield, J. J. 205, 215
Blum, T. 210, 215
Bodrova, M. B. 113, 134
Boelter, M. D. 133
Boenheim, F. 130, 132
Boggs, E. M. 84
Bohatyrtscuk, F. 170, 195
Boling, J. L. 107, 132
Bourdillon, R. B. 107, 132
Boyes-Watson, J. 37, 84
Bozler, E. 262, 263, 268
Braumer, L. 128, 132
Breder, C. M., Jr. 112, 132, 133
Brenizer, A. G., Jr. 118, 132
Brice, A. T. 137, 140
Brice, B. A. 84
Briggs, L. J. 104, 132
Britten, R. H. 215
Bronk, D. W. 261, 262, 268
Brosteaux, J. 75, 84
Brown, A. 21, 22, 37, 38
Brown, M. V. 114–116, 132, 133, 135
Brücke, E. 130, 132
Brues, A. M. 213, 215
Bruner, F. H. 210, 211, 215
Buchthal, F. 132, 248, 249, 251, 253, 260, 268
Buhmann, A. 132
Bull, H. B. 24, 38, 96, 97, 103, 132, 246, 248, 253, 268
Bullock, T. H. 105, 132
Bunker, J. W. M. 212, 214
Bünning, E. 128, 132
Burch, C. R. 137, 140
Burch, G. J. 113, 133
Burge, E. L. 106, 132
Burge, W. E. 106, 132
Burr, H. S. 104–108, 110, 132, 134
Burrows, H. 110, 132
Bushey, M. S. 136
Butler, J. A. V. 30, 38

Cameron, W. H. 206, 208, 215
Campbell, A. B. 132
Campbell, D. H. 84
Campbell, W. W. 133
Canti, R. G. 138, 140
Canton, W. 134
Carmosino, J. A. 106, 135
Carr, C. W. 136
Caspersson, T. 261, 268
Castle, W. B. 204, 215
Cavallo, T. 86, 132

Cavendish, H. 86, 87, 132
Cecil, R. 14, 15, 18, 23, 24, 26, 37, 38
Chagas, C. 112, 132
Chapeau, M. 132
Cholodny, N. G. 105, 132
Claessons, S. 146, 163
Clark, G. L. 170, 195
Coates, C. W. 111, 112, 114–116, 124, 132, 133, 135
Cohn, E. J. 15, 17, 37, 44, 71
Cohn, W. E. 118, 132, 133, 200, 214
Cole, K. S. 89, 104, 123, 133
Compton, A. H. 166, 167, 195
Conlon, P. 204, 210, 215
Cope, O. 118, 132
Copp, D. H. 214, 215
Copson, E. T. 163
Cook, G. H. 146, 163
Coryell, C. D. 111, 132
Coulson, C. A. 14, 37, 163
Cox, J. T. 163
Cox, R. T. 86, 87, 111, 112, 114–116, 124, 132, 133, 135
Cox, S. T. 14, 37
Crane, E. E. 108, 109, 116, 117, 119–121, 123, 124, 126–130, 133
Craver, L. F. 209, 215
Craxford, S. R. 100, 133
Crook, A. W. 152, 163
Crowfoot, D. 23, 38
Crowther, J. A. 170, 196
Curtis, B. R. 89, 134
Curtis, H. J. 133
Cushing, H. 86, 133

Dainty, M. 266, 268
Dalgiesh, E. 71
Damon, E. B. 133
Darmon, S. E. 64, 71, 72
Darsie, M. L. 124, 132
Dauvillier, A. 169, 195
Davidson, E. 37, 84
Davies, R. E. 97, 116, 117, 119–121, 123–125, 128–130, 133
Davis, H. 106, 107, 135
Davson, H. 117, 133
Dean, R. B. 97, 133
Debye, P. 80, 84
Delrue, G. 119, 133
Dent, R. F. 103, 135
Dervichian, D. G. 84
Dickinson, S. 72
Doan, C. A. 211, 215
Donnan, F. G. 5, 37, 97, 116, 133
Donné, A. 87, 133
Doty, P. M. 26, 31, 38, 84, 235, 267
Drinker, C. K. 204, 215
Drinker, K. R. 204, 215
Du Bois-Reymond, E. 87, 108, 114, 133
Duke-Elder, W. S. 117, 133
Du Mesnil de Rochester, R. 229, 241, 267
Dunlap, C. E. 210–212, 215

Edelhoch, H. L. 77, 84
Edsall, J. T. 11, 15, 22, 37, 38, 44, 71, 77, 84
Edser, E. 163
Einstein, A. 9, 37, 77, 84
Elder, J. H. 107, 133
Elliot, A. 72
Emerson, R. 133
Enelow, A. J. 119, 135
Engelhardt, W. A. 266, 268
Engelmann, T. W. 130, 133
Engström, A. 170, 173, 181, 187, 194, 196, 261, 268
Eriksson-Quensel, I. B. 163
Ettisch, G. 87, 133
Evans, R. D. 204–212, 214, 215
Evers, J. 104, 105, 128, 133
Ewart, J. C. 112, 113, 133

Fåhraeus, R. 16, 17, 37
Failla, G. 207, 215
Fankuchen, I. 72, 83, 84
Faraday, M. 87, 133
Faure-Fremiet, E. 241, 268
Fell, H. B. 139, 141
Feng, T. P. 261, 268
Fenn, W. O. 255, 268
Ferri, C. 225, 267
Fessard, A. 112, 115, 131, 133
Finch, G. 107, 133
Finkel, M. 213, 215
Fischer, E. 247, 268
Fletcher, C. M. 248, 263, 268
Fletcher, E. S. 136
Flexner, L. B. 102, 117, 133, 136
Flinn, F. B. 208, 215
Flint, L. H. 132
Flory, P. J. 28, 38
Floyd, W. F. 133
Folk, F. 134
Foort, L. 228, 267
Forkner, C. E. 211, 215
Foster, J. F. 22, 38
Fowler, R. H. 27, 28, 38
Fournet, G. 84
Fraenkel, G. 72
Francis, W. L. 124, 126, 133
Friedenwald, J. S. 102, 117, 133

FRIEDGOOD, R. B. 106, 107, 135
FRITSCH, G. 112, 113, 133
FRUTON, J. S. 116, 133
FULTON, J. H. 86, 133

GALLAGHER, D. .M 215
GALVANI, L. 86
GASSER, H. S. 256, 257, 259, 268
GATTY, O. 97, 100, 124, 133
GEE, G. 31, 38
GERARD, R. W. 99, 133, 265, 268
GETTLER, A. O. 204, 215
GILSON, A. S. 266, 268
GLASSTONE, S. 103, 133
GOBY, P. 169, 195
GODDARD, D. R. 133
GOODMAN, E. N. 119, 131, 133, 135
GORDON, M. A. 206, 215
GORIN, M. H. 103, 131
GOSTING, L. J. 14, 37, 163
GOTCH, F. 112, 113, 133
GOUY, A. 145, 163
GRACE, E. J. 211, 215
GRAHAM, J. 99, 133
GRALEN, N. 37
GRANT, W. M. 134
GREEN, D. E. 133
GREEN, H. S. 84
GREENBERG, D. M. 118, 133
GREGOR, H. P. 136
GROSS, D. 18, 37
GRUNDFEST, H. 89, 133
GUERIN, M. 213, 215
GUGGENHEIM, E. A. 5, 8, 27–29, 31, 37, 38
GUINIER, A. 80, 84
GUNTER, G. 134
GUTFREUND, H. 16, 18–21, 23, 24, 26, 34, 37
GUTTMAN, R. 123, 133

HALL, C. E. 72, 247, 268
HALLIMOND, A. F. 154, 163
HAMILTON, J. G. 200, 214, 215
HAMMETT, F. S. 105, 132
HANSZEN, A. H. 105, 134
HARMAN, P. J. 106, 133
HARRIS, M. 228, 267
HARRIS, R. S. 212, 214, 215
HARTLEY, G. S. 33, 38
HARTMANN, H. 128, 134
HARTREE, W. 248, 261, 268
HARVEY, S. C. 132
HASELBACH, C. 228, 267
HEYCOCK, C. T. 169, 195
HEALD, R. H. 132
HEILBRUNN, L. V. 88, 99, 108, 113, 134
HELLER, W. 84
HENDREY, J. H. 17, 37
HENLY, W. 86, 134
HENSELEIT, K. 121, 134
HERBERT, D. 83, 84
HERDAN, G. 13, 37
HERZOG, R. O. 20, 37
HESS, V. F. 205, 206, 215
HILDEBRAND, J. H. 27–29, 38
HILL, A. V. 248, 255–259, 261, 262, 268
HILL, R. T. 132
HINSHELWOOD, C. W. 111, 134
HITCHCOCK, D. T. 101, 103, 134
HOBER, R. 96, 99, 106, 108, 117, 133, 134
HODGKIN, A. L. 99, 117, 134
HOFER, A. W. 140, 141
HOFFMAN, F. L. 204, 215
HOGBEN, L. T. 260, 268
HOKIN, L. E. 117, 119, 124, 135
HOLLANDER, F. 116, 134
HOLM JENSEN, I. 118, 134
HØNCKE, P. 255, 268
HOPPE, W. 193, 196
HOVLAND, C. I. 105, 132
HOYT, R. C. 105, 134
HUF, E. 117, 124, 134
HUGGINS, M. L. 28, 38, 49, 59, 72
HUGHES, A. F. W. 138–141
HUMPHREY, G. F. 257, 268
HUMPHRIES, R. E. 211, 215
HUNTER, J. 86, 112, 134
HUXLEY, A. F. 117, 134

IBALL, J. 132
INGALLS, E. N. 19, 20, 37
IVES, J. E. 206, 208, 215

JAKUS, M. A. 72, 247, 268
JOHNSON, P. 14, 37
JOHNSON, W. A. 134
JOHNSTON, J. P. 15, 24, 37, 38
JÖNSSON, E. 167, 195
JORDAN, H. J. 264, 268
JORDAN, W. K. 84, 266, 268
JORGENSEN, C. B. 118, 134
JOSEPH, N. R. 102, 134
JULLANDER, J. 14, 37
JUPNIK, H. 140

KALCKAR, M. 133
KAISER, E. 248, 249, 251, 260, 268

Kaplan, E. 134
Katz, B. 99, 117, 134, 258, 264, 268
Kaye, M. 239, 267
Keele, C. A. 133
Kegeles, G. 14, 37, 163
Kekwick, R. A. 21, 37
Keller, R. 130, 134
Kelly, N. B. 106, 132
Kendall, J. 5, 37, 38
Kendrew, J. C. 20, 37, 38
Keynes, R. D. 99, 117, 118, 134
Keys, A. 117, 134
Khalil, F. 121, 134
Kinder, W. 151, 153, 163
Kinsley, V. E. 117, 118, 134
Kirkpatrick, P. 169, 195
Kirkwood, J. G. 84, 233, 267
Klaus, K. 108, 134
Klein, O. 166, 195
Kleinzeller, A. 268
Klevens, H. G. 84
Klopstock, E. 136
Knappeis, G. G. 248, 249, 251, 261, 268
Knef, J. P. 204, 210, 215
Knowles, E. L. 205, 215
Koch, K. 128, 134
Kogan, D. A. 134
Köhler, A. 137, 140
Komagata, S. 103, 134
Korr, I. M. 101, 134
Krajuchin, B. V. 113, 134
Kratky, O. 84
Krebs, H. A. 121, 134
Krogh, A. 100, 116, 117, 134
Kuffler, S. W. 264, 265, 268
Kumanomido, 134
Kümmel, 134
Kurzrok, R. 102, 134

Lacassagne, A. 205, 210, 215
Lamarque, P. 170, 195
Lamm, O. 18, 37, 153, 155, 163
Langman, L. 107, 134
Laporte, Y. 265, 268
Lattes, J. S. 205, 215
Lawrence, A. S. C. 268
Lea, D. 26, 38
Leaderman, H. 227, 267
Leatherdale, D. 170, 196
Levi, H. 118, 134
Levin, A. 257, 268
Lewis, W. H. 138, 140
Lindström, B. 173, 196
Linfoot, E. H. 155, 163
Linnik, W. 151, 152, 163
Lipmann, F. 116, 134
Lipmann, G. 87, 134
Lisco, H. 213, 215
Livingood, J. J. 134
Ljubimova, M. N. 266, 268
Lloyd, D. M. 239, 267
Loeb, J. 134
Lohmann, K. 135
Lomax, R. 72
Lomholt, S. 205, 215
Longmuir, N. M. 116, 117, 119, 120, 124, 128, 133
Longsworth, L. G. 13, 14, 37, 163
Longuet-Higgins, C. H. 16, 37
Lonti, R. 77, 84
Loos, W. 137, 140
Ludford, R. J. 139, 141
Lund, E. J. 101, 105, 124, 126, 133–135
Lundegårdh, H. 117, 134, 135
Luneberg, R. K. 158, 163
Luthy, H. 181, 196

Machado, A. L. 124, 135
Macinnes, D. A. 163
Macleod, J. 135
Macmahon, H. E. 213, 125
Macpherson, C. F. C. 13, 37
Maddigan, S. E. 170, 195
Mahan, R. I. 105, 134
Marceau, F. 240, 267
Markham, R. 26, 38
Marmont, G. 264, 268
Marsh, B. S. 255, 268
Marshall, C. 89
Martin, A. J. P. 72
Martin, L. C. 137, 140
Martland, H. S. 204, 209–211, 215
Marwick, T. C. 72
Matthews, B. H. C. 87, 88, 135
Maurice, D. M. 117, 133
Maxwell, J. C. 87, 135
McNiff, W. T. 205, 206, 215
Meader, R. G. 89
Meier, R. 135
Meitina, R. A. 266, 268
Menkin, M. F. 107, 108, 135
Mercer, E. H. 229, 267
Merton, T. 150, 163
Meyer, K. H. 99, 135, 217, 218, 222, 225, 228, 229, 232, 247, 248, 267, 268
Meyerhof, O. 133, 135, 254, 256, 268
Miall, M. 268
Michaelis, L. 95, 99, 135
Michel, K. 137, 140
Miller, A. R. 28, 31, 38
Miller, G. L. 23, 38
Miller, J. R. 103, 135
Minot, A. S. 205, 215

Mislowitzer, E. 119, 135
Mizell, L. R. 228, 267
Mizushima, S. 72
Mogila, M. T. 113, 134
Mommaerts, W. F. H. M. 266, 268
Mond, R. 119, 135
Mongar, J. L. 234, 267
Moore, D. H. 13, 37
Morris, J. P. 207. 215
Morrison, P. R. 77, 84
Mottram, J. C. 210, 215
Moullin, E. B. 135
Moyer, L. S. 103, 131, 132
Mudd, S. 103, 130, 135
Muller, F. H. 219, 267
Munch-Peterson, A. 266, 268
Murayama, M. M. 133
Murphy, A. S. 213, 215
Musselman, L. K. 106, 107, 132

Nachamsohn, D. 124, 125, 135
Needham, D. M. 268
Needham, J. 268
Nerancy, J. T. 207, 215
Neville, F. H. 169, 195
Nicholas, J. S. 106, 135
Nichols, J. B. 17, 37
Nieuwenhoven, L. M. 248, 264, 268
Nishina, Y. 166, 195
Nord, F. F. 84
Norris, C. 204, 215
Northrop, F. S. C. 104, 132

Oakley, H. B. 14, 37
Ockendon, F. E. J. 163
Oettlé, A. G. 140, 141
Ogston, A. G. 8, 12, 14–16, 18, 23, 24, 26, 37, 38, 95, 97, 99, 100, 116, 125, 133, 135, 145, 163
Oncley, J. L. 10, 18, 22, 37, 38
Onsager, L. 36, 38
Oster, G. 4, 26, 37, 38, 77, 84, 266, 268
Osterburg, H. 137, 140
Osterhout, W. J. V. 99, 100, 135
Overbeek, J. T. G. 36, 38

Palmer, A. H. 24
Pardee, A. B. 84
Parker, H. M. 201, 214
Parmenter, R. 107, 135
Patterson, W. B. 119, 135
Payne, B. O. 137, 140
Pedersen, K. O. 14–16, 18, 23, 24, 37, 38, 163
Peelman, I. 200, 214
Pelc, S. R. 137, 140
Perry, S. V. 72, 232, 267
Perutz, M. F. 18, 19, 25, 37, 38, 54, 72, 84
Peterfi, T. 87, 132, 133
Petit, J. L. 259, 268
Philpot, J. St. L. 14, 37, 146, 155,163
Picken, L. E. R. 229, 248, 267, 268
Pickett, M. J. 102, 132
Pinhey, K. F. 260, 268
Pledge, H. T. 87, 135
Polson, A. 18, 22, 37, 38
Ponder, E. 135
Porod, G. 84
Porter, R. R. 20, 37
Postma, N. 264, 268
Prentiss, A. M. 17, 37
Preston, M. M. 139, 141
Pringle, J. W. S. 265, 268
Pumphrey, R. 121, 124, 133, 135
Putzeys, P. 75, 84

Quigley, J. P. 119, 135

Ramsey, R. W. 231, 248, 252, 255, 258, 262, 267, 268
Ransmeier, R. E. 265, 268
Rayleigh, Lord 74, 84, 145, 155, 163
Reboul, J. 106, 107, 135
Reed, C. I. 134
Reed, R. 72, 232, 267
Rehm, W. S. 110, 117, 119, 124, 126, 128–130, 135
Reitter, G. S. 210, 215
Renk, F. 248, 268
Rice, H. V. 117, 119, 124, 135
Richards, O. W. 139–141
Riley, D. P. 79, 83, 84
Riseman, J. 233, 267
Ritchie, J. M. 265, 268
Robertson, R. N. 117, 120, 135
Robinson, M. E. 21, 37
Roche, A. 37
Roche, J. 37
Rock, J. 107, 108, 135
Roe, E. M. F. 110, 132
Rogers, P. V. 106, 135
Rosenblith, W. A. 116, 132
Rosene, H. F. 99, 104, 124, 135
Rosenhagen, H. 134
Rosenthal, M. 211, 215
Ross, R. T. 117, 119, 124, 135
Rothschild, M. 119, 135
Rothschild, Lord 104, 135

ROTHENBERG, M. A. 124, 135
ROUSSEY, G. 213, 215
RUDALL, K. M. 72, 229, 267

SABIN, F. R. 211, 215
SACKS, J. 118, 135
SANDOW, A. 265, 268
SANGER, F. 20, 24, 37, 38
SANKEWITSCH, E. C. 105, 132
SARRE, H. 119, 135
SCATCHARD, G. 21, 22, 32, 37, 38
SCHELKUNOFF, S. A. 163
SCHLUNDT, H. 207–209, 215
SCHMIDT, C. L. A. 15, 37
SCHMIDT, W. J. 238, 261, 267
SCHMITT, F. O. 247, 268
SCHOBER, B. 110, 132
SCHOEPFLE, G. M. 266, 268
SCHRANK, A. R. 128, 135
SCHULZ, G. V. 28, 30, 34, 38
SEABORG, G. T. 200, 214
SEEDS, W. E. 84
SEEMAN, H. E. 171, 196
SEEMAN, J. 257, 268
SEIDLIN, S. M. 208, 215
SEIL, H. A. 206, 208, 215
SEKORA, A. 84
SELBIE, F. R. 213, 215
SENTI, F. R. 24, 25, 38
SHEN, S. 267, 268
SICHEL, F. J. M. 252, 268
SIEGBAHN, M. 167, 195
SIEVERT, R. 170, 171, 195
SILVER, S. 119, 135
SJÖGREN, B. 22, 23, 38
SKOW, R. K. 124, 132
SMILES, J. 139, 141
SMITH, G. M. 110, 132
SMITH, K. M. 26, 38
SMITHIES, O. 116, 125, 135
SMYTHE, C. V. 133
SNODGRASS, J. 107, 108, 135
SNYDER, I. W. 217, 267
SÖRENSON, S. P. L. 24
SPARK, L. C. 72
SPEAKMAN, J. B. 231, 267
SPEISER, R. 84
SPRINGER, S. 112, 132
SPROULL, W. T. 170, 195
STAPP, P. 126, 134, 136
STECK, I. E. 134
STEINBACH, H. B. 97, 136
STETTEN J. R. DE W., Jr. 119, 135
STIEHLER, R. D. 102, 117, 133, 136
STOCK, J. P. 137, 140
STOKES, A. R. 84
STREET, A. 71
STREET, S. F. 252, 255, 268
STRONG, L. C. 110, 132
SUTHERLAND, G. B. B. M. 64, 71, 72
SVEDBERG, T. 1, 11, 12, 14–18, 22, 23, 37, 38
SVENSSON, H. 145, 155, 163
SWANN, M. M. 138–140
SWYNGEDAUW, R. 119, 136
SZENT GYÖRGYI, A. 232, 266, 267

TAFFEL, M. 132
TAYLOR, E. W. 137, 139–141
TEMPLE, R. B. 72
TEORELL, T. 100, 117, 119, 133, 136
TERNER, C. 130, 136
THEORELL, H. 19, 20, 37
THIMANN, K. V. 105, 128, 136
THOMAS, H. E. 210, 211, 215
THORELL, B. 261, 268
THOVERT, J. 155, 163
TIBBETS, D. M. 215
TISELIUS, A. 12, 18, 37, 153, 163
TITAEV, A. A. 119, 136
TRILLAT, J. J. 171, 196
TRISTRAM, G. R. 37
TRURNIT, H. J. 193, 196
TUCKETT, R. F. 235, 267
TURCHINI, J. 170, 195
TWYMAN, F. 163

UMRATH, K. 102, 136
USSING, H. H. 99, 118, 134, 136

VAN DER WYK, A. J. A. 218, 222, 267
VAN SLYKE, D. D. 97, 136
VENCHIKOV, A. I. 119, 136
VERWEY, E. J. W. 36, 38
VICTOREEN, J. A. 166, 195
VIOL, C. H. 206, 208, 215
VISSCHER, M. B. 118, 136
VOLTA, A. 86, 136
VON ARDENNE, M. 171, 196
VON HAMOS, L. 192–194, 196

WALLS, G. L. 113, 136
WALKER, W. C. 86, 136
WALLER, A. D. 87, 136
WALSH, J. 112, 136
WARNER, R. C. 24, 25, 38
WASSERMAN, A. 234, 267
WATSON, G. N. 163

WAUGH, D. J. 23, 38
WEBER, H. H. 84, 232, 267
WEIL, P. E. 210, 215
WELCH, F. V. 139, 141
WENT, F. W. 105, 128, 136
WERSÄLL, R. 117, 119, 136
WHEWELL, C. S. 51, 72, 238, 267
WIEGAND, W. B. 217, 267
WIENER, O. 9, 37
WIERSMA, C. A. G. 264, 268
WIGGLESWORTH, V. B. 117, 136
WILKINS, M. H. F. 84
WILKINS, M. J. 117, 135
WINTON, F. R. 235, 249, 253, 263, 267
WOELFFLIN, R. 241, 268
WOHLGEMUTH, J. 136
WÖHLISCH, E. 217, 229, 241, 246, 248, 253, 267, 268
WOOD, R. W. 163
WOODS, H. J. 51, 71, 72, 226–228, 230, 238, 267
WYMAN, J. 18–20, 37, 257, 268

YERKES, R. M. 107, 133

ZEEMAN, P. 152, 163
ZERNIKE, F. 137, 140, 145, 146, 148, 149, 153–155, 163
ZIMM, B. H. 26, 28, 30, 38, 84
ZOLLINGER, H. V. 138–141

SUBJECT INDEX

Absorption of x-rays in biological materials 167 *et seq*
— studies 117
Acetylcholine, synthesis of 125
Acetylgalactosamine 60
Acetylglucosamine 60
Actin 57 *et seq*
Adductor of *Venus verrucosa* 240
Adenosine triphosphate, action of on muscle proteins 266, 267
Afocal methods of refractometry 144 *et seq*
Albumin, egg 45, 53
Alginate fibres, contraction in 234
Alginic acid 61, 62, 65
— fibres 62
— — and pectin 61
α-keratin 56, 228
— transformation from feather keratin 57
α-type chain 47
Amino acids in polypeptide chains 41
Amplifocal methods of refractometry 150 *t seq*
Anaesthesia 106
Analysis, cytochemical, 164
Aperfocal methods of refractometry 153 *et seq*
— — — — efficiency and linearity of 161 *et seq*
Arthropods 60, 62, 64
Autoradiographs 138

Backbone linkage 43
β-lactoglobulin 24
Biocolloids, physical constants of 25 *et seq*
Bioelectric potentials, maintenance and function 85 *et seq*
Biological material, assay of 164 *et seq*
— systems, fundamental structures in 39 *et seq*
Bone necrosis 210
Bushy stunt virus 25

Calcium in dentine cells 190
Carcinogenesis 110
Catfish, electric 113
Cell structures, mass of, from x-ray investigations 173 *et seq*
Cellulose 61, 62
Chemical engines 234
— — model of 235 *et seq*
Chironomus salivary gland 182
— giant chromosome 183
Chitin 60 *et seq*
— of arthropods 62, 64
— — *Phycomyces* 64
— — *Sirex gigas* (ovipositor) 64
— purified 61
Cholinesterase 124
Close packing of protein molecules 45
Clupein thymonucleate 67, 68
Coherence, of light beams 156
Collagen-type structures 58 *et seq*
— — beef tendon fibrils 59, 60
— — cartilage ground substance 59
Collagen-type structures, cuticle of annelids 60
— — cuticle of nematodes 60
— — egg capsules of skate 60
— — elastoidin 59
— — filaments ejected by sea cucumber 60
— — ithyocol of swim bladders 59
— — mesogloea substance of jellyfish 59
— — shell membrane of bird's egg 60
— — tendons of vertebrates 59
— — white connective tissue fibres 59
Complex amplitude, definition of 157
Corpuscular proteins 54
Cultures of vertebrate tissues 138
Cuticle of annelids 60
— — arthropods 60
— — earthworm 70, 71
— — nematodes 60
Cystine bridges 44
— — epidermal structures in birds and reptiles 57
— — keratin, in birds and reptiles 55
— — — in seagull quill 56
Cytochemical analysis 164
Cytochemistry of cells 164

Denaturation of proteins 54
Dentine cells, calcium and phosphorus in 190
Diffusion 2, 9, 14
— potentials 96
Donnan equilibrium 6, 32, 36, 97

Earthworm cuticle, 70, 71
Edestin 45
Egg albumin 45, 53
Elastic power cycles 231 *et seq*
— strains, thermodynamics of 216 *et seq*
Electrical energy, utilization of 127
Electric eel 111
— fish 86
— organs 110, 113, 115
— power of tissues 125
Electrokinetic potentials 102
Electron microscope 3, 66
— — and beef tendon fibril 59
— — and earthworm cuticle 70
— — and thymonucleic acid 67
— transfer 101
Electrophoresis 2, 12
Electroplax 89, 113
Energy relations in gastric mucosa 129
Entropy calculations 217
— cycles and heat engines 233
— of dilution 28
— of mixing 27, 33
Epidermal structures of birds and reptiles 57
Epidermin 47
Epithelium, ectodermal, in annelids 60
— skin 183

Feather keratin structures 55 *et seq*
Fertilization of rat egg 139
Fibres, alginate 234

Fibres and muscles, mechanical properties of 216 *et seq*
— hair 227, 246
— muscle 181, 246 *et seq*
— plastic 236
— and tendons 238 *et seq*
Fibrils in crossed lamella systems 70
Film emulsions 172
Fission, nuclear 197
— products 213, 214
Focal methods of refractometry 146
Forces, intermolecular 31, 69
Frictional ratio, calculation of 10
Frozen rubbers, energy relations of 229

Gastric mucosa 116, 119
— — energy relations of 129
— — of frog 98, 127
Glycogens 26
Gouy's method of refractometry 145
Growth, effect of electric current on 104 *et seq*
Gymnotus electricus, discharge of 86

Haemoglobin 12, 16, 31, 32, 34
Hair and wool 47, 227
— human, elastic properties of 227
— — energy relations of 230
Healing of wounds 108, 109
Helix buccal retractor 262
Hyaluronic acid 65
Hydrochloric acid, secretion of in gastric mucosa 116, 120
Hydrogen bonds and infra-red studies 43

Influenza virus 81
Infra-red absorption 43
Insulin 22, 53
— fibrous 23
Interferometer, Michelson 151
Intermolecular forces 31, 69
Isotopes, radioactive 117, 118, 197 *et seq*
— — half lives of 201

Keratin 44, 45
— in reptiles and birds 55

Lactoglobulin 35
Lamm scale method of refractometry 145
Large molecules, properties of solutions of 1 *et seq*
Latency relaxation 265
Light scattering and osmotic pressure 78
— — — size of proteins 75
— — by bovine serum albumin 78
— — — bushy stunt virus in water 75
— — — influenza virus in water 81
Light scattering by monodispersed system 76
— — — polydispersed system 76
— — — protein solutions 73 *et seq*
— — — small independent particles 74 *et seq*
— — — small non-independent particles 76 *et seq*
— — — tobacco mosaic virus solutions 77
— — experimental arrangements 75
Liquid junctions, diffusion potentials 96

Macromolecules, properties of solutions of 1 *et seq*
Mass absorption coefficient 166, 186, 187
— distribution in cells 181
Membrane permeability 99
— potentials 95, 97, 98
Membranes, electrical characteristics of 122
— secretory 118 *et seq*
Mesodermal cells 60
Metabolism 123
— of radium 204 *et seq*
Methaemoglobin 54
Michelson interferometer 151
Microcuvettes, zinc in 190
Microradiography 169
— techniques 170
Microscope, electron 3, 59 *et seq*
— Zernike 137
Microscopy, narrow pencil illumination 144, 145
— phase contrast 137 *et seq*
— ultraviolet 139
Mobilities of ions 96
Modulus, dynamic 236, 239, 240, 249
— of elasticity 222
Molecular weight, calculation of 10
Monolayers 45
Mouse spleen culture 139
Muscle 246 *et seq*
— and adenosine triphosphate 253, 256
— as diphase system 253
— as engine 216
— bioelectric potentials 87
— clam 257
— contractions 263
— dogfish jaw 257
— dynamic modulus of 249
— elastic properties of 246
— fibres 181
— frog sartorius 259
— helix 262
— jaw, of dog fish 257
— lantern, of *Echinus* 257
— latency relaxation 265
— *Mytilus* anterior byssal retractor 249, 262
— optical changes 247
— temperature effects and heat exchanges 247
— tonus 261
— viscous-elastic properties 255
Myoglobin 19
Myosin 26, 44, 45, 47
Mytilus anterior byssal retractor 249, 262
— — — — stimulation of 264

NERVE, action currents 87
— cell 182
— fibres 180
Neutrons, effects of 197
Nucleic acids 66 *et seq*
Nucleolus 182
Nucleotides 26
Nylon, energy relations of 230

ONION root tip 182
Optical density 75, 81
Orientation phenomena 69 *et seq*
— — and fibrils in crossed lamella 70
Oscillograph 88
Osmotic activity 35
— pressure 2, 4, 14, 27 *et seq*
— — and light scattering 78
— — of haemoglobin 34
— — — proteins 31
— — — serum albumin 34, 35
Ovalbumin 20
— x-ray scattering 83
Ovulation 106, 107, 108
Oxidation potentials 101

PECTIN 61
Phase contrast microscopy 137 *et seq*, 154
— — — and bacteriophage 140
— — — — fertilization of rat egg 139
— — — — mitosis 139
— — — — mitochondria 140
— — — of chromosomes 138
— — — — malignant tumours 139
— — — — tissue-culture cells 138
Phosphate bonds, high energy 116
Phosphorus in dentine cells 190
Pile, atomic 197
Plant growth, effect of electric current on 128
Plutonium 239 197
Polyalanine 61
Polyglycine 61
Polymers, high 31
Polynucleotides 66 *et seq*
Polypeptide chains, structure of extended 41 *et seq*
— — backbone linkage between 43
— — crystallite 42
— — in hair and wool 48
— — interactions between side chain groups 43
— — x-ray spacings in 41
Polysaccharides, straight chain 61 *et seq*
Polyvinyl chloride, elastic properties of 226, 227
Potential energy cycles 232
Potentials, bioelectric 85 *et seq*
— — and physiological changes 104 *et seq*
— — as source of energy 123
— oxidation-reduction, 101–102
— phase boundary 100
Protein chains, unfolding of in hair and wool 48
Proteins, axial ratios 10
Proteins, close packing of 45
— fibrous and corpuscular 54 *et seq*
— homogeneity of 11
— molecular constants of 15 *et seq*
— molecules, close packing of 45
— size of 75
Psittacosis and *Vaccinia* virus 140

RADIATION safety factor 199
Radioactive elements, decay chains 203
— isotopes, half lives of 201
— — tolerance of man for 197 *et seq*
— — use of 117, 118
— paint 204
Radium, deposition of 202
— elimination of 208, 209
— metabolism 204 *et seq*
— retention in body 206, 207, 208, 210
— toxic concentrations 212
Raoult's law 29
Rat egg, fertilization and early development of 139
Ray, electric 112
Rayleigh refractometer 146
Rayleigh's law 74
Refractometry 142 *et seq*
— afocal methods of 144 *et seq*
— amplifocal methods of 150 *et seq*
— aperfocal methods of 153 *et seq*
— focal methods of 146 *et seq*
— Gouy's method 145
— Lamm scale method 145
Reptiles, epidermal structures of 57
— feather keratin structures in 55 *et seq*
Rubber like solids, elasticity of 217
Rubbers, energy relations of 229
— frozen 225
— imperfect 224

SAFETY factor for radiation, 199
Scattering 73 *et seq*
— by large independent particles 78 *et seq*
— — large non-independent particles 81 *et seq*
— — small independent particles 74 *et seq*
— — small non-independent particles 76 *et seq*
— of light 2
— — — by proteins 73
— of x-rays by proteins 73
Schlieren method of focusing 155
Secretion of hydrogen chloride 116, 120
— studies 117
Secretory function 116 *et seq*
— membranes 120–122
Sedimentation velocity 2, 8
Serum albumin 21, 31, 32 *et seq*, 78
Side chain packing in polypeptide chains 44
Single nerve fibres 180
Sirex gigas, chitin of ovipositors of 64
Skin, sulphur in 190, 191
Sodium thymonucleate films 66
Soft x-rays 164 *et seq*

Solids, characteristics of normal and rubbery 216
Spherical crystal lenses 193
Structures in biological systems 39 *et seq*
Supercontraction phenomena 50 *et seq*
Svedberg units 8
Swim bladder of perch 121

TEMPERATURE effects and heat changes 247
Tendons 238 *et seq*
— relaxation and 'set' of 245
Thermodynamic analysis of real solutions 27
Thymonucleic acid 67
Thymus gland of mouse 68
Tolerance criteria 199, 202 *et seq*
— dose of radium 199
— — — radioactive isotopes 198
— — — x-rays 199
Torpedo 110, 112
Transformations between corpuscular and fibrous conditions in proteins 54
Tropomyosin 45
Tumour, malignant 139
Turbidity 75

ULTRACENTRIFUGE 8
— optical methods for 146
Unit cell of chitin 63
Uranium 235 197
— — decay chain 203
— — fission products 200

VAN DER WAALS FORCES 43
Vertebrate tissues, cultures of 138
Virus particles 25
— tobacco mosaic 25, 36
— tomato bushy stunt 25
Viscosity 2
— effect of on chains 223
Viscous-elastic effects 224, 241, 242
— properties 235
Voltaic pile 86

WOUNDS, healing of 108, 109
— potential differences in 109

X-RAY, absorption and scattering o 165
— analysis of crystals 144
— diffraction by actin 57
— — — alginic acid 62, 65
— — — cellulose 62
— — — chitin 62, 64
— — — collagen-type structures 58
— — — egg albumin 53
— — — epidermin 47, 51, 52
— — — feather keratin 55, 56
— — — fibrin 51
— — — fibroin 44
— — — fibrous and corpuscular proteins 54
— — — insulin 53
— — — keratin 44, 51, 56
— — — myosin 44, 47, 52
— — — polynucleotides 66, 67
— — — polypeptide chains, α-type 46
— — — — — β-type 52
— — — — — side-chain packing 44
— — — — — in hair and wool 44
— — — polysaccharides 61
— — — proteins 4, 16
— emission analysis 191, 195
— image formation 169
— microradiography, techniques 170
— microscope 169
— scattering by actomyosin solutions 80
— — — edestin solutions 80
— — — haemocyanin solutions 80
— — — haemoglobin solutions 80
— — — insulin solutions 80
— — — ovalbumin solutions 80, 83
— — — protein solutions 73 *et seq*
— — — tobacco mosaic virus solutions 80
— tube, diagram of 177
x-rays, soft 164 *et seq*
— absorption of in biological material 167
— spectrography 164, 184
— emission analysis of biological material 191

ZERNIKE MICROSCOPE 137
Zeta potential 103

PREFACE

THE EDITORS have had some difficulty in deciding what is the proper field to be covered by reviews of recent progress in biophysics. Excluding biochemistry on the one hand and physiology on the other, there lies between a vast and rather amorphous field of study of which the frontiers and lines of demarcation are anything but well defined. On the one hand there is the application of the methods of physics to the study of the characteristic molecules of life—proteins, nucleoproteins, lipoproteins and other complex molecules of many and mostly unexplored varieties, a study which should perhaps be strictly regarded as an aspect of biophysical chemistry—hence the sub-title of the book. On the other hand there is the application of physical methods and instruments to the study and elucidation of living structures such as nerve and muscle fibres, a field in which modern refinements offer the greatest promise of progress in the future; progress which will perhaps bring its subjects into the first category. Both these fields of work are represented in the present volume and the editors have endeavoured to maintain a reasonable balance between them.

It has been our aim to obtain readable and interesting reviews of recent progress in selected subjects. Our contributors have been asked not merely to compile a précis of recent papers, or annotated lists of references, but to write critical reviews in which they discuss the subjects from their own point of view; which may be read with profit by many who are not experts and which will provide scientists with a general survey of recent work and ideas.

Since it is impossible in a single volume to deal at length with more than a few topics, it is intended that this volume will be followed by others at regular intervals. In time a more complete coverage of all the varied aspects of this rapidly developing branch of science will thus be built up.

One omission from this volume possibly calls for comment *viz* the absence of any article on the use of radioactive isotopes as tracers in biology. While we are not unaware of the great potential usefulness of this method we felt that the studies which have so far been made have been very fully reviewed in other publications and that a further review at the present time would be unlikely to contain significant new material. We hope that as it develops both the occasion and opportunity of including reviews of the progress of this method will arise.

May, 1950

J. A. V. BUTLER
J. T. RANDALL

ACKNOWLEDGMENTS

THE AUTHORS and publishers wish to thank the following authorities, and also the authors concerned, for permission to reproduce diagrams.

American Association for the Advancement of Science :	*Figures 1*, p 75 ; *5*, p 81
American Physiological Society :	*Figures 5a*, p 120 ; *8*, p 129
Biochemical Journal :	*Figures 5b*, p 120 ; *7*, p 127
Chemical Review :	*Figure 3*, p 79
Elsevier Publishing Co. Inc. :	*Figures 15*, p 56 ; *7*, p 83
Journal of General Physiology :	*Figures 3*, p 114 ; *4*, p 115
Nederlandsche Chemische Vereniging :	*Figure 6*, p 82
New York Zoological Society :	*Figures 3*, p 114 ; *4*, p 115
University of Texas Press :	*Figure 1*, p 105

1

PROPERTIES OF SOLUTIONS OF LARGE MOLECULES

H. Gutfreund

LARGE molecules or macromolecules are usually defined as molecules of molecular weight greater than 10,000. Probably the two most important differences between solutions of large molecules and ordinary solutions are, first, the magnitude and direction of the deviation from the ideal solution laws and secondly, the methods to be employed for the study of thermodynamic and kinetic properties of such solutions. Only properties of solutions of large molecules of biological origin will be discussed here. The study of such systems and especially of solutions of physiologically active proteins has become increasingly important as more and more enzymes, hormones, respiratory pigments, immunological antigens and antibodies and plant and animal viruses have been prepared in a highly purified and crystalline form. It is essential to obtain detailed information about the molecular constants (size, shape, hydration, surface charge and activity) of these substances under varied conditions of solution before it is possible to attempt the interpretation of the chemical and physical mechanism of the reactions which these substances catalyse or which they undergo.

It is the purpose of the present article to discuss some of the special methods applicable to studies on solutions of large molecules and to give a theoretical analysis of the data obtained. The resulting information can be divided into three groups :

1 physical constants of individual molecules
2 aggregation and disaggregation reactions
3 the influence of inter- and intramolecular forces on the properties of solutions.

Since the first correct values for the molecular weight of some proteins were obtained by ADAIR (G. S.)[1] in the years 1920–22, by the osmotic pressure method, and somewhat later by SVEDBERG from ultracentrifugal sedimentation and diffusion measurements, these and other methods for the determination of molecular constants of proteins have been developed to such an extent that, for homogeneous systems, they have become very accurate and almost a routine procedure. One cannot, however, emphasize sufficiently often how important it is to use many different methods on the same solution of macromolecules, as only when good agreement among them is

reached, can there be real confidence in results obtained on such complicated systems.

When the stage is reached that the molecular constants of a particular protein, or other macromolecular substance, are accurately known, further studies on solutions of varied solute concentration, solvent composition and temperature can give valuable information about intramolecular forces, holding subunits of large molecules together, and intermolecular or long range forces causing interaction between solute molecules or solute and solvent molecules.

METHODS FOR STUDYING SOLUTIONS OF LARGE MOLECULES

A large number of methods for the determination of physical constants of macromolecules is now in use, and it is worth while enumerating all those available, together with the properties upon which they depend.

The molecular weight or size of the solute molecule is the most important property and holds a coordinating position among its physical constants as all these depend, at least to some extent, on the molecular dimensions. This will be seen clearly when the experimental methods are listed together with the properties upon which the results obtained depend.

1 Osmotic pressure and membrane potential *2* Pressure of surface film *3* Sedimentation equilibrium	depend upon the activity of the solute, and give the molecular weight, and charge, in the case of membrane potential, in ideal or very dilute solutions.
4 Sedimentation velocity *5* Diffusion constant	depend upon molecular size and frictional properties (shape and solvation) if data from both methods are available these properties can be evaluated.
6 Viscosity and other hydrodynamic methods	depend upon shape and molecular volume (solvation).
7 Dipole moments and relaxation constant	depend upon shape and charge distribution.
8 Electrophoresis	depends upon surface charge and frictional properties, the latter to the same extent as sedimentation and diffusion.
9 Scattering of light and x-rays	depends upon molecular size and shape.

10	Crystallographic unit cell measurements by x-ray diffraction	applied to highly ordered systems can give information about molecular size, shape and solvation.
11	Electron microscope	can be used for direct measurements of molecular dimensions in the dry state if these are greater than 100 Å.
12	Ultra filtration through graded membranes	can give approximate molecular dimensions.
13	Chemical analysis	depends upon minimum equivalent molecular weight.
14	Titration curves	depend upon charges per mol. equivalent.

The homogeneity of the solute is usually investigated by examination in the ultracentrifuge or the electrophoresis apparatus or by solubility studies. Careful analysis of diffusion curves also yields information on that point. More detailed attention will be given to this problem later on.

It is, of course, impossible to do more than mention most of the methods of the above list. The choice of those described in some detail in the following sections (osmotic pressure, diffusion and ultracentrifugal sedimentation measurements) may appear somewhat arbitrary but studies with these techniques have supplied us with the bulk of the reliable constants of large molecules of biological origin. Results of osmotic pressure measurements also give data for the free energy of the solution which already allow some interesting thermodynamic interpretations and promise more detailed information about the structure of solutions and intermolecular forces. The osmotic pressure procedure has many advantages over any of the other methods yielding thermodynamic information as it is technically much simpler than light scattering or sedimentation equilibrium measurements and considerably more accurate than vapour pressure or freezing point determinations. As for the measurements of surface film pressures, these are only strictly applicable if adsorption is reversible and this is not the case for most proteins, which are the most important compounds among large molecules of biological origin. It should however be added that the light scattering technique, which is in a very active stage of development, both as regards experimental procedure and theoretical interpretation, promises to be very useful. This method can give data on solutions of particles of much greater molecular

weight ($M >$ 1,000,000) than could be studied usefully by osmotic pressure, furthermore it is almost independent of the distribution of small electrolytes. It can also be used to determine rates of changes, such as disaggregation or aggregation phenomena taking place in solution, as it gives instantaneous results. A number of reviews on this subject have recently been published and references to these, as well as to original papers on the lines mentioned above, can be found in OSTER's article[2].

OSMOTIC PRESSURE OF SOLUTIONS

In a system where the solvent is separated from a solution by a membrane which is permeable to the solvent but not permeable to the solute, the osmotic pressure is defined as the pressure to be applied to a solution to bring it into equilibrium with the pure solvent. At constant pressure the partial molar free energy of the pure solvent is greater than that of the solvent in the solution, and the excess pressure (osmotic pressure) needed to equalize the partial molar free energy of the solvent in the two phases is a measure of the free energy of dilution:

i.e. $$\partial G = -\partial PV_o \qquad \ldots.(1)$$

The osmotic pressure equation for a binary mixture composed of solvent and solute, can be derived as follows : n_1 is the number of mol of solvent and n_2 the number of mol of solute.

The partial free energy of dilution of the solvent is

$$\Delta\bar{G}_1 = \partial \Delta G/\partial n_1 \qquad \ldots.(2)$$

The partial heat of dilution of the solvent is

$$\Delta\bar{H}_1 = \partial \Delta H/\partial n_1 \qquad \ldots.(3)$$

The partial entropy of dilution of the solvent is

$$\Delta\bar{S}_1 = \partial \Delta S/\partial n_1 \qquad \ldots.(4)$$

The partial molar volume of the solvent is

$$\bar{V}_1 = \partial V/\partial n_1 \qquad \ldots.(5)$$

where V is the volume of the solution. These equations can be inserted in

$$\Delta\bar{G}_1 = \Delta\bar{H}_1 - T\Delta\bar{S}_1 \qquad \ldots.(6)$$

A solution is called ideal or perfect if the partial molar free energy of the solvent, and with that the osmotic pressure and vapour pressure, is proportional to the mol fraction N_1 of the olvent over the whole range of concentrations from $N_1 = 1$ to $N_1 = 0$, where

$$N_1 = n_1/(n_1 + n_2) \qquad \ldots.(7)$$

There is no heat or volume change on dilution of an ideal solution:

i.e. $$\Delta\bar{H} = 0 \qquad \ldots.(8)$$

$$\Delta V_1 = \bar{V}_1 - V_1 = 0 \qquad \ldots.(9)$$

where V_1 is the volume of one mol of pure solvent. It follows therefore that the partial molar free energy of dilution of an ideal solution is

$$\Delta\bar{G}_1 = -T\Delta\bar{S}_1 \qquad \ldots.(10)$$

It can be shown by statistical mechanics that for a regular solution the partial molar entropy of dilution of the solvent

$$\Delta\bar{S}_1 = -R \ln N_1 \qquad \ldots.(11)$$

The derivation of equation 11 and definition of the term regular solution will be found in a later section. It follows from equations 10 and 11 that

$$\Delta\bar{G}_1 = RT \ln N_1 \qquad \ldots.(12)$$

and from equations 1 and 12 that the osmotic pressure

$$P = -RT/V_1 . \ln N_1 \qquad \ldots.(13)$$

in very dilute solutions

$$-\ln N_1 = -\ln (1 - N_2) = +N_2 = n_2/n_1 \qquad \ldots.(14)$$

and equations 13 and 14 give

$$P = RTn_2/n_1V_1 \qquad \ldots.(15)$$

and if $n_1\bar{V}_1 = V$ then $P = RTn_2/V$ $\qquad \ldots.(16)$

Equation 16 which is only valid for extremely dilute solutions, is identical with van't Hoff's osmotic pressure equation, but it is derived from thermodynamic principles and not from the misleading analogy of 'gas pressure' soundly criticized by KENDALL[4] and GUGGENHEIM[5] (see also DONNAN[6] for a discussion of various derivations of the osmotic pressure equation). Kendall emphasizes that all the components of a liquid solution are in the liquid state, and it would therefore be very difficult to obtain correct expressions from kinetic theory, especially for deviations from the ideal laws. It must be possible to formulate a kinetic picture of osmotic pressure, but as long as there is no certainty about the correct physical mechanism of a process it is clearly more logical to use a derivation which makes no assumptions about it. The arguments about a suitable model for the kinetic process causing osmotic pressure may go on for a long time, but the results derived from thermodynamic reasoning present an indisputable answer. This clearly shows the advantages and disadvantages of thermodynamic methods in general; accurate results can be obtained without the use of any assumptions about the mechanism of a process but no information about physical occurrences is given.

As osmotic pressure is a measure of the free energy of dilution (see equations 1 and 12) and the free energy is the sum of the negative entropy and the heat change, deviations from the ideal osmotic pressure law can be due to anomalous entropy changes and/or due to

heat of interaction. The causes of the deviations from the ideal solution laws will be discussed later on. Before such an examination can be carried out, the ideal properties of the particular solution under investigation must first be ascertained. The molecular weight of the macromolecular solute can be evaluated from the results of osmotic pressure measurements, if these are carried out on solutions over a range of solute concentrations (C in gm of solute per litre of solvent) and P/C is plotted against C. From the extrapolated value for P/C at $C = .0$ correct values for the solute molecular weight can be calculated from equation 16.

$$P = RTn_2/V = RTC/M$$

Hence

$$M = RTC/P \qquad \ldots\ldots(17)$$

It has previously been stated that very dilute solutions behave as ideal solutions. Once the molecular weight of the solute has been determined accurately the magnitude of the deviation of the partial molar free energy of mixing from the ideal entropy of mixing can be calculated from the experimental data. It will be seen that much useful information about the structure of solutions and properties of large molecules, besides molecular weight data, can be obtained from osmotic pressure measurements.

The Donnan equilibrium— The above treatment of osmotic pressure only considered solutions of molecules of zero overall surface charge, no matter whether they contain no charged groups or an equal number of positive and negative groups. The osmotic pressure of charged macromolecules is the sum of the colloid osmotic pressure and the osmotic pressure due to the unequal distribution of transferable ions, which will occur in such systems.

The following derivation for the distribution of ions across a membrane separating a solution of protein in dilute hydrochloric acid from its dialysate will give a suitable demonstration. If a large volume of dialysate is used its composition is constant and

$$[H^+]_{Dial} = [Cl^-]_{Dial}$$

The following two conditions enable one to set up equations which describe the distribution of ions across the membrane. Electric neutrality in the solutions requires that if the molarity of the protein is m and the charge valence $\mathcal{Z}^+$, the concentration of $[Cl^-]_{Sol}$ ions in the solution

$$[Cl^-]_{Sol} = m\mathcal{Z}^+ + [H^+]_{Sol}$$

The second, the Gibbs-Donnan equilibrium distribution, requires that the products of the diffusible ion concentrations on both sides of the membrane are equal. At equilibrium the transfer of ∂n sodium

and chloride ions from one side of the system to the other, at constant temperature and volume, should cause no free energy change. No electrical work is being done as an equal number of positive and negative ions are being transferred, hence

$$\partial nRT \ln \frac{[H^+]_{Sol}}{[H^+]_{Dial}} + \partial nRT \ln \frac{[Cl^-]_{Sol}}{[Cl^-]_{Dial}} = 0$$

Hence

$$[H^+]_{Sol} \times [Cl^-]_{Sol} = [H^+]_{Dial} \times [Cl^-]_{Dial}$$

To simplify the writing let us take

$$[Cl^-]_{Sol} = y \qquad [H^+]_{Sol} = x$$

and

$$[H^+]_{Dial} = [Cl^-]_{Dial} = a$$

Then

$$y = x + m\mathcal{Z} \qquad \ldots\ldots(18)$$

and

$$x \times y = a^2 \qquad \ldots\ldots(19)$$

from equations 18 and 19 it follows that

$$a = \sqrt{x(m\mathcal{Z} + x)} \qquad \ldots\ldots(20)$$

and the ion concentration difference

$$\begin{aligned} \Delta &= x + y - 2a \\ &= m\mathcal{Z} + 2x - 2\sqrt{x(m\mathcal{Z} + x)} \end{aligned} \qquad \ldots\ldots(21)$$

The sum of the colloid osmotic pressure P_p and the pressure P_i due to excess electrolytes

$$P_{obs} = P_p + P_i = mRT + (m\mathcal{Z} + 2x - 2\sqrt{x(m\mathcal{Z} + x)})RT$$

If a solution of colloidal electrolytes is dialysed against pure water

$$a = 0 \qquad x = 0 \qquad y = m\mathcal{Z}$$

the osmotic pressure is dependent on the colloid charge concentration

$$P = m(\mathcal{Z} + 1)RT \qquad \ldots\ldots(22)$$

The partial pressure due to the Donnan effect can be reduced by increasing the diffusible ion concentration. If equation 21 is used together with the expression

$$x = \tfrac{1}{2}(-m\mathcal{Z} + \sqrt{(m\mathcal{Z})^2 + 4a^2})$$

obtained from equations 18 and 19, it can be shown that Δ decreases as $a/m\mathcal{Z}$ increases. If the composition of the dialysate, and with that a, is known, the value for x can be determined electrometrically or from chemical analysis. From these data and the above equations the excess ion pressure P_i and the charge valence $\mathcal{Z}$ of the colloid can be calculated. The suppression of the Donnan effect through increase of the concentration of small ions is of great value for the measurement

of properties of solutions of colloidal electrolytes. The osmotic pressure effects due to the unequal distribution of small electrolytes can be much greater than the actual colloid osmotic pressure. These phenomena are of importance in the interpretation of thermodynamic equilibria in biological systems and also in the stability of some gels in vitro.

The present discussion neglects corrections by means of activity coefficients but it is sufficiently accurate for the understanding of the phenomena to be described later on. Guggenheim and more recently Ogston[3] criticized the use of electrometric methods for the determination of membrane potentials on thermodynamic grounds. This may invalidate the accurate quantitative interpretation of the data but does not detract from the practical importance of the influence of ' Donnan effects ' on the properties of solutions of colloidal electrolytes.

ULTRACENTRIFUGAL SEDIMENTATION

The sedimentation velocity of a particle in a centrifugal field depends upon the gravitational force acting on the particle

$$\phi(\rho_p - \rho)\,\omega^2 x$$

and upon the frictional resistance, which, according to Stokes' law for solid spherical particles in very dilute solutions, is

$$6\pi r\eta\,.\,dx/dt$$

Steady rate of sedimentation is reached when the centrifugal force is equal to the frictional resistance. The centrifugal force per mol is

$$M(1 - V\rho)\,\omega^2 x$$

and the frictional resistance per mol (take $6\pi r\eta = F$)

$$\mathcal{N}F\,.\,dx/dt$$

and if

$$M(1 - V\rho)\,\omega^2 x = \mathcal{N}F\,.\,dx/dt$$

$$\frac{dx/dt}{\omega^2 x} = \frac{M(1 - V\rho)}{\mathcal{N}F} = S \qquad \ldots\ldots(23)$$

The sedimentation constant S is usually expressed in Svedberg units : 10^{-13} cm/sec/dyne. It can be calculated from the measurements of movements of the sedimenting boundary. The symbols used in the above expressions are defined as :

ϕ = volume of one particle
ρ_p = density of particles
ρ = density of medium
ω = angular velocity of centrifuge
x = distance from centre of rotation
r = radius of particles
η = viscosity of medium
dx/dt = velocity of particle
M = molecular weight of particle
V = partial specific volume of particle
$\mathcal{N}$ = Avogadro's number

Diffusion— The discussion of diffusion velocity will be limited here to the treatment of particles which are either uncharged or accompanied by sufficiently high concentrations of small electrolytes so that the large molecules behave as uncharged particles.

The diffusion constant D is defined by Fick's law, which states that the quantity of substance diffusing across a boundary of unit area in unit time is equal to the product of D times the concentration gradient

$$q = -D.dc/dx \qquad \ldots.(24)$$

This equation can be used for the practical evaluation of D if an analytical procedure is used to determine the amount of solute diffusing across a solvent/solution boundary.

Often it is very useful, for a detailed analysis of diffusion phenomena, to obtain the data in the form of a plot of the concentration gradient against distance across the boundary. In the ideal case this plot will be in the form of a Gaussian probability curve, as has been shown by WIENER[7] in 1893. Wiener has also demonstrated optical methods for observing boundaries and recording distribution curves. The second moment (or the square of the standard deviation) of such probability curves correspond to Δ^2, the mean square displacement in EINSTEIN'S[8] diffusion equation

$$D = \Delta^2/2t \qquad \ldots.(25)$$

Expressions 24 and 25 can be used for the determination of diffusion constants from experimental data. It remains now to show some relation between the diffusion constant and the properties of the particles in solution. The force exerted by the molecules diffusing across a boundary of concentration gradient dc/dx can be expressed as an osmotic force. The force f exerted by one particle in a very dilute solution is given by

$$f = \frac{RT}{cN}\frac{dc}{dx} \qquad \ldots.(26)$$

The frictional force $F.dx/dt$ will govern the rate of steady diffusion

$$\frac{dx}{dt} = \frac{RT}{cN} \cdot \frac{dc}{dx} \cdot \frac{1}{F} \qquad \ldots.(27)$$

It can be seen from equation 24 that

$$D = \frac{dx}{dt} \cdot c \Big/ \frac{dc}{dx} \qquad \ldots.(28)$$

and from equations 27 and 28 it follows that

$$D = RT/NF \qquad \ldots.(29)$$

These derivations are applicable only to solutions of slowly diffusing

substances, where the height of the solvent and solution columns can be regarded as infinite.

CALCULATION OF MOLECULAR WEIGHT AND FRICTIONAL RATIO

The molecular weight of large molecules can be calculated from equation 30, which is derived from 29 and 23

$$M = RTS/D(1 - V\rho) \qquad \ldots.(30)$$

The effect of frictional resistance is eliminated in the above expression, but nevertheless it is only applicable to results of measurements on very dilute solutions. The procedure to obtain accurate values for molecular weights from sedimentation and diffusion measurements is analogous to that followed in the case of osmotic pressure. A series of measurements of D and S over a range of solute concentrations must be carried out and values obtained from extrapolations to zero concentration should be used for the calculation of M.

As already mentioned, Stokes' empirical law states that the frictional resistance offered to a spherical particle of radius r moving in a medium of viscosity η at a velocity dx/dt is $6\pi r\eta dx/dt$ (or $F.dx/dt$). If the sedimenting or diffusing molecules are solvated, F will be increased as it will correspond to the radius of the solvated molecules. If the molecules are asymmetric F will also be increased as it will correspond to the radius of the spherical volume swept by the rotating molecules. Much use is being made of the term frictional ratio (f/f_o) in the descriptions of properties of large molecules in solution. This is the ratio of the frictional force acting on the actual particle in solution to the force acting on a non-solvated sphere of the same molecular weight. If the molecular weight of a substance is calculated from the experimental values for D and S, the diffusion and sedimentation constants D_o and S_o of solid spherical particles of the same molecular weight can be derived. It can be seen readily that

$$f/f_o = D_o/D = S_o/S \qquad \ldots.(31)$$

Unfortunately in many discussions of the shape of protein molecules the effect of solvation on the frictional ratio has been neglected and absurdly high values for the axial ratio a/b are sometimes recorded in the literature. Generally proteins and other large molecules are divided into two groups : globular and fibrous, though some have the shape of rigid rods and few of the so called globular ones are really spherical. As it is difficult to determine either shape or hydration independently of each other, information about these two properties is therefore still somewhat scanty. ONCLEY[9] prepared some very valuable graphs showing the simultaneous influence of hydration and

asymmetry on the frictional ratio and the viscosity coefficient. It is of some interest to note that a spherical particle with a solvation of 0·3 gm per gram of dry solute has the same frictional properties as an unsolvated particle of axial ratio a/b about 3·3. It will be seen later on, when the constants of some well defined proteins are being discussed, that the hydration of some of the globular ones is about 0·3 gm of water per gram of protein. x-ray diffraction and light scattering data promise to give independent information on the shape, and it will be seen that the combined data obtained for some proteins agree well with the frictional ratio. The method of measuring streaming birefringence developed by EDSALL[10] also gives independent information about the shape of large molecules in solutions.

HOMOGENEITY OF PROTEINS

Proteins are the only macromolecular substances which are anything like homogeneous and this section will therefore be devoted solely to them. SVEDBERG's studies in the ultracentrifuge, which enabled him to observe moving boundaries in centrifugal fields equivalent to about 300,000 times gravity, showed that proteins are molecular entities and probably monodisperse or very nearly so.

A survey of the current literature shows that ultracentrifugal sedimentation, electrophoresis and solubility studies are the most common criteria for homogeneity employed for the examination of samples of proteins. The concentration gradient distribution curves, obtained if one of the refractive index methods is used for the observation of diffusion, can also give some information on the homogeneity of the diffusing solute. The experimentally observed diffusion curve is superimposed over an ideal Gaussian distribution curve with the same second moment ; if the solute is homogeneous with respect to diffusion velocity, the two curves are identical.

When a solvent/solute boundary moving in an electric or gravitational field is studied, it is possible to get one of three results

1 one homogeneous boundary, the spread being due to diffusion only
2 several individual boundaries
3 one boundary which is spreading more rapidly than can be accounted for by diffusion.

It is evident that a careful examination of the homogeneity of a protein sample by one of the moving boundary methods requires also an accurate knowledge of the rate of diffusion.

As each of the methods mentioned above will only test for homogeneity with respect to one property, it is essential to have satisfactory results from all of them before it can be accepted that a sample is

really homogeneous. When SVEDBERG and TISELIUS developed the moving boundary methods of observing sedimentation and electrophoresis, respectively, it became fairly widely accepted, for a time, that the appearance of a single boundary on examination of a sample with either of these techniques, was sufficient for it to qualify as homogeneous. During the last few years however, more rigorous tests were applied to many of the purified or even crystalline proteins, and at the moment a wave of scepticism is flowing through the literature as one by one they are being found to be heterogeneous.

Examples are known in which electrophoresis did not show impurities in protein samples which could be detected in the ultracentrifuge or from solubility studies. In spite of this, conscientious electrophoretic examination in solutions of varied *p*H is probably the most sensitive single test available for homogeneity. The electrophoretic investigations followed two lines. The control of fractionation procedures by examination in the Tiselius apparatus at each stage ; this has been applied mainly to those proteins which could be shown to resolve into separate boundaries during migration in the electric field. ALBERTY[11], who followed another line, namely the study of boundary spreading, writes

> The usual criterion for the electrophoretic homogeneity of a protein is that it migrates as a single boundary in an electric field in buffers of various ionic strengths and hydrogen concentrations. However, this alone is not sufficient evidence that all the protein molecules have the same electrophoretic mobility and further evidence regarding the electrophoretic homogeneity may be obtained by studying the rate with which the protein gradient spreads in the electric field.

Alberty describes a method for determining the standard deviation from the mean mobility and ANDERSON and ALBERTY[12] applied this procedure to fourteen purified proteins, seven of them crystalline. None of the preparations satisfied their criteria for purity. There are of course quite a number of physical phenomena which can give rise to excessive boundary spreading during electrophoresis, but it has been pointed out by OGSTON[13] that no other known phenomenon besides heterogeneity will cause uniform reversible spreading of both the ascending and descending boundary. This criterion has been employed by Alberty and his collaborators. It can readily be shown that the difference of one valence is sufficient to cause resolution into two boundaries ; a very good demonstration of this can be given if a mixture of haemoglobin and oxyhaemoglobin or other reduced and oxygenated haem proteins is examined. It would therefore not be surprising if all the proteins which are shown to be heterogeneous from boundary spreading could be resolved into separate fractions at a suitable *p*H.

The homogeneity of well defined crystalline proteins is only just

beginning to be investigated. It remains to be seen whether proteins are essentially heterogeneous or whether further fractionation will give products which satisfy our criteria for purity. It is quite possible that the instability of proteins causes the observed heterogeneities and the very sensitive electrophoretic test described above, as well as further theoretical investigations may result in the discovery of physical phenomena for the observed spreading. Just because it has been shown *e.g.* in serum albumin and ovalbumin, that even crystalline proteins could be subjected to further fractionation, one cannot hope to improve all the proteins by this procedure. That instability is the cause of heterogeneity at least in some proteins has been shown by MACPHERSON, MOORE and LONGSWORTH[14].

This description of the problem of the homogeneity of proteins gives only a rough outline of a not very satisfactory state of affairs. Fortunately from the point of view of most physical examinations, there are a series of proteins which can justifiably be interpreted as homogeneous systems in solution. In an analysis of homogeneity with respect to size a sense of proportion must be kept between the accuracy of the methods involved and the heterogeneities encountered. This particular criterion of size is however more applicable in the case of industrial high polymers and polysaccharides which are known to be heterogeneous with respect to size and in which the size distribution is of great interest.

The mean values obtained from molecular weight measurements on heterogeneous systems depend of course, on the method used. Of the two methods which produced most of the data discussed in the following sections, osmotic pressure measurements give the number average molecular weight,

$$M_N = \frac{\Sigma M_i n_i}{\Sigma n_i} \qquad \ldots\ldots(32)$$

and combined diffusion and sedimentation constant give the weight average molecular weight

$$M_W = \frac{\Sigma M_i g_i}{\Sigma g_i} = \frac{\Sigma M^2_i n_i}{\Sigma M_i n_i} \qquad \ldots\ldots(33)$$

Comparison of the two average values obtained gives a measure of the molecular weight distribution. HERDAN's[15] recent note describes how the standard deviation from the mean molecular weight can be derived directly from the above two equations 32 and 33.

EXPERIMENTAL METHODS AND THEIR ACCURACY

Apparatus will not be described in any detail but reference ought to be made to sources where a full working knowledge of the various

procedures for measuring osmotic pressures, diffusion constants and sedimentation constants can be obtained. The development of methods for the accurate measurements of osmotic pressures of solutions of large molecules is mainly due to ADAIR, whose various types of osmometers are described in a recent textbook[16] and to OAKLEY[17]. Many modifications have been made by various workers and one or two different types have also been developed ; reference to them will be found in ALEXANDER and JOHNSON's book[16]. Attention should be drawn to the recent development of an osmotic balance by SVEDBERG and JULLANDER[18]. Though very elaborate, the osmotic balance is much more accurate than any of the other methods for measuring the osmotic pressure and promises to be very useful in the study of high molecular weight or very dilute systems.

Most of the cells for diffusion measurements are described by Alexander and Johnson. A very valuable new contribution has been made by COULSON, COX, OGSTON and PHILPOT[19] who developed a new optical system and cell which enables one to measure diffusion constants much more rapidly, on more dilute solutions and with greater accuracy than has hitherto been possible. An optical system similar to theirs has also been described by KEGELES and GOSTING[20] and tested experimentally by LONGSWORTH[21].

Three types of ultracentrifuges are in use, oil turbine, air turbine and electrically driven ones. Though some more recent models have been constructed with certain modifications, the descriptions given in SVEDBERG and PEDERSEN's book[22] are quite adequate.

The accuracy of the results of osmotic pressure measurements depends on the error involved in measuring the solute concentration and the pressure. The former can always be estimated with a maximum error of $\pm$ 1 per cent. If suitable methods are being employed the osmotic pressure can be determined with the same accuracy. With the simplest osmometers the pressure can be measured to $\pm$ 0·1 cm of water and with refined osmometers and a cathetometer the pressure can be measured to $\pm$ 0·01 cm. The osmotic balance should be able to measure to $\pm$ 0·001 cm and can therefore be used to determine pressures of about 0·1 cm with sufficient accuracy. Unfortunately it is difficult to obtain such precision in measurements on protein solutions, for besides the difficulties always encountered in solutions of colloidal electrolytes, instability of these substances as well as the tendency to adsorb irreversibly on to the membranes, can cause considerable disturbance of osmotic equilibria.

The accuracy of the measurement of sedimentation constants has been fully discussed in a recent paper by CECIL and OGSTON[23]. They point out that most of the errors in published data are due to faulty

speed and temperature measurements. The very valuable and thorough investigations of Cecil and Ogston are worthy of the most careful study by anybody who intends to determine sedimentation constants. Neither sedimentation nor diffusion constants should be published unless the standard deviation from the mean of a fair number of measurements is not more than one per cent. As will be seen in the following section it takes a large number of years to displace wrong data which have found their place in the tables of textbooks. Both sedimentation and diffusion constants should always be measured over a range of concentrations and extrapolated to zero concentration. The convention of quoting D_{20} and S_{20} as extrapolated diffusion and sedimentation constant in water at 20°C should be adhered to. It will be shown in the following section that really careful evaluation of constants, even in such labile systems as solutions of proteins, are rewarded by reliable results. That is, under such circumstances molecular constants obtained from methods involving entirely independent experimental procedure and theoretical bases are in close agreement.

MOLECULAR CONSTANTS OF SOME PROTEINS

It would be of little value to reproduce the long lists of proteins and their physical constants which have been published in textbooks such as SVEDBERG and PEDERSEN[22], SCHMIDT[24] and COHN and EDSALL[25]. Neither is it the purpose of the present section to collect data which are the result of single investigations on newly isolated or not previously examined proteins ; information about this latter group is contained in the articles on proteins and amino-acids in the *Annual Review of Biochemistry*. It is thought to be more important within the scope of this article, to consider the constants of a few crystalline proteins which have been studied by several workers by a number of different methods. The pioneer work of Svedberg and his collaborators contributed a very large amount of useful information during the first two decades of the life of the ultracentrifuge. The time has now come, however, when very much more accurate results of measurements in solutions are needed in conjunction with chemical and crystallographic data, to help in elucidating the structure of proteins or the mechanisms of their functions.

The theory of Svedberg that the molecular weights of proteins are multiples of 17,000 fits some of the respiratory proteins, but it has been shown by chemical analysis and end group assay that the subunits in various haemoglobins and myoglobins differ in their chemical composition[26], and the theory of fixed multiple molecular weights does not appear to be significant even for this group of proteins. JOHNSTON,

LONGUET-HIGGINS and OGSTON[27] have shown the scheme to be untenable for all proteins on statistical grounds, even using the table of data published by Svedberg and Pedersen. More up to date values for some of the molecular weights produce an even more marked spread of the groups of values.

Most of the information made use of during the discussion in this and the following sections was obtained from osmotic pressure, diffusion and sedimentation measurements. For some proteins very accurate information about the molecular weights and dimensions are available from x-ray diffraction measurements on single crystals and we shall make use of some data obtained from such investigations. A recent statement[28] that ' the presence of impurities can give rise to completely erroneous results if x-ray analysis is used for the estimation of molecular weight or dimensions ' needs a considerable amount of qualification. If the impurity remains amorphous during crystallization of the main component (this is the most probable thing to happen) measurements on the crystals of the latter will not be affected at all by the presence of the former. Mixed crystals can always be detected by a competent observer, and for the formation of single crystals containing both main component and impurity in a crystalline form, the two proteins would have to be so much alike as not to affect the measurements. It is worth noting in this connection that even such related proteins as adult and foetal haemoglobin will not crystallize together[29]. It is very likely that crystallographic methods can yield more accurate data about the dimensions of protein molecules than any other procedure.

Haemoglobin— Certainly no other protein has been studied as extensively as haemoglobin by all the physical methods available. The reason for this, besides its physiological importance, is the ease with which large quantities of this protein can be prepared in fairly pure and crystalline form.

ADAIR (G. S.)[1, 30] carried out a number of investigations on the osmotic pressure of solutions of haemoglobin from various sources and found that for all those examined the molecular weight was about 67,000. GUTFREUND[31] has included more recent measurements in a summary of all the molecular weight data from osmotic pressure investigations on mammalian haemoglobins and pointed out that for all those investigated in any detail the molecular weight, shape and hydration, are the same within the limits of the experimental accuracy. They differ however, in many of their other properties (surface charge, solubility, physiological activity and chemical composition). Shortly after Adair's osmotic pressure measurements gave the first correct results for the molecular weight of haemglobin, confirmation of his findings was obtained by SVEDBERG and FÅHRAEUS[32] during

investigations with the ultracentrifuge. The iron content of haemoglobin (0·34 per cent) gave a very precise estimate of the minimum molecular weight (16,700) (for a summary see COHN, HENDREY and PRENTISS[33]) and so even after the early work of Adair and Svedberg it could be accepted that the molecular weight was $4 \times 16{,}700 = 66{,}700$ and was accurately known. This fact made haemoglobin a very useful model for the study of thermodynamic and kinetic properties of protein solutions. These will concern us more in a later section, but during the osmotic investigations of Adair and others[30, 31] the value of 67,000 for the molecular weight of many mammalian haemoglobins was confirmed whenever extrapolation to zero concentration was used to demonstrate the behaviour of ideal solutions of these proteins. Horse, human and sheep haemoglobins have been studied by crystallographic methods[29] and the values for the molecular weight obtained were in excellent agreement with that given above.

Unfortunately the published results of sedimentation and diffusion measurements on haemoglobins are not nearly as reliable and comprehensive as the osmotic work. This is probably due to the fact that there has been no real doubt about the accurate value for the molecular weight of haemoglobin since 1926, while the repeated and more accurate investigations of sedimentation and diffusion constants which were carried out on other proteins were called for by continuous arguments over the various values calculated for the molecular weights. SVEDBERG and FÅHRAEUS's[32] figure of 67,000 for the molecular weight of horse CO-haemoglobin was calculated from sedimentation equilibrium measurements. This method has great theoretical advantages over both osmotic pressure or sedimentation and diffusion measurements as it can give simultaneous information about homogeneity, molecular weight and weight distribution. Furthermore, it gives information about the solvent activity over a range of concentrations during one experiment, and is based on thermodynamic theory. No description of this method has been given in this article, since due to great practical difficulties the accuracy of results obtained so far compares unfavourably with that of osmotic pressure or sedimentation and diffusion measurements. The first sedimentation velocity measurements on horse haemoglobin were carried out by SVEDBERG and NICHOLS[34]. This investigation is of great historical interest, but the optical, temperature and speed measurements used could not yield data which satisfy the demands on precision made today. Furthermore the invaluable practice of quoting figures for D_{20} and S_{20} corrected to diffusion and sedimentation in water at 20°C only came into use fairly recently, and it is difficult to compare some of the older values. The only figures for sedimentation constants of haemoglobins quoted

in the tables of SVEDBERG and PEDERSEN[22] are unpublished observations on horse and human haemoglobins. The sedimentation constant is stated to be the same for both proteins $S_{20}=4{\cdot}4\times10^{-13}$ cm/sec.

Cecil and Gutfreund examined samples of horse, sheep and human haemoglobin in the ultracentrifuge and concluded that their sedimentation constants were identical ($S_{20} = 4{\cdot}5 \times 10^{-13}$ cm/sec), but these values do not compare in accuracy with some of the thorough examinations of sedimentation constants carried out recently[23]. Renewed ultracentrifugal studies of different haemoglobins are needed now that it has been shown by OGSTON and his collaborators[23, 35] how reliable results can be obtained by that method.

If the molecular weight $M = 66{,}700$ and the sedimentation constant $S_{20} = 4{\cdot}5 \times 10^{-13}$ the frictional ratio $f/f_o = 1{\cdot}15$ (equation 31). If this value for the frictional constant is used to derive an estimate of the hydration and asymmetry of haemoglobin molecules with the help of ONCLEY's[9] graphs, it is found that one combination of these properties resulting in $f/f_o = 1{\cdot}15$ is a hydration of 0·3 gm of water per gram of protein and an axial ratio of 1·7. This is in excellent agreement with data obtained from crystallographic measurements by PERUTZ and his collaborators[29, 36]. It is encouraging to find such agreement between data for hydration and shape obtained from x-ray measurements on wet crystals and kinetic measurements in solutions. Perutz suggests that the horse haemoglobin molecule is in the form of a cylinder with the two parameters 34 Å and 57 Å respectively. It would be premature, however, to argue from that that the shape and hydration of protein molecules in crystals and in solution are always the same. WYMAN[37] finds that the dimensions found for haemoglobin molecules from x-ray and centrifugal studies accord well with his own findings from dielectric dispersion.

No really accurate data for the diffusion constants of any haemoglobins are available. TISELIUS and GROSS[38] studied horse haemoglobin and report a value of $D_{20} = 6{\cdot}3 \times 10^{-7}$ sq cm/sec, while LAMM and POLSON[39] found $D_{20} = 6{\cdot}9 \times 10^{-7}$ sq cm/sec for human haemoglobin. Theoretically the diffusion constant of haemoglobin can be calculated from equation 30; if $M = 67{,}000$ and $S = 4{\cdot}5 \times 10^{-13}$ cm/sec it follows that $D = 6{\cdot}48 \times 10^{-7}$ sq cm/sec. In both experimental investigations referred to above, it was found that the diffusion constant increased with dilution below a protein concentration of 0·5 per cent. From this it was concluded that horse and human haemoglobin molecules tend to dissociate in dilute solutions. Confirmation of this phenomenon has been obtained by Pedersen from ultracentrifugal investigations. The evidence about the dissociation of haemoglobin in solution is, however, at present somewhat confusing. There is some

indication that the tendency to dissociate varies with the source from which the haemoglobin is taken. We have found that sheep CO-haemoglobin underwent no change in dispersion on dilution from 0·5 to 0·15 per cent. The sedimentation constants found for this protein at the two concentrations were $4{\cdot}4 \times 10^{-13}$ and $4{\cdot}5 \times 10^{-13}$ respectively. A small increase of the sedimentation constant with decreasing concentration is of course to be expected in solutions where the molecular weight is independent of concentration (pp. 10, 15). Besides this dependence on species it is also known that the solvent composition influences the dissociation of haemoglobin molecules (GUTFREUND[31]).

It is evident from the above description that the molecular weight of mammalian haemoglobins is accurately known and that the shape and hydration of these molecules is known to a good approximation. However, a considerable amount of further useful information about haemoglobin molecules will be obtained if the recent improvements in the accuracy of sedimentation and diffusion studies are applied to haemoglobin solutions under varied conditions. It may well be that some of the conflicting views on the dissociation of haemoglobins (ADAIR (G. S.)[40]) are due to the instability of haemoglobin. Either the formation of met-haemoglobin or the ' physiological inactivation ' which occurs in oxyhaemoglobin could cause a change in the stability of the globin molecules ; samples of haemoglobin during purification, crystallization or storage often lose their power to combine reversibly with oxygen. Some structural changes occur in haemoglobin during the reduction (PERUTZ, unpublished observation) and if the haem is split off from haemoglobin in some cases the resulting globin molecule dissociates into smaller units[41, 42].

Myoglobin— It is quite logical that those who have been interested in the structure of haemoglobin should turn their attention to myoglobin. From a crystallographic point of view it seemed particularly promising to study protein molecules which have a similar physiological function but only a quarter of the size of haemoglobin molecules. Only horse muscle myoglobin will be discussed here.

THEORELL[43], who isolated this protein, established that its iron content is the same as that of haemoglobin and its minimum molecular weight is therefore 16,700. For some years there was a considerable amount of conflicting evidence as to whether the molecular weight was 17,000 or 34,000. Myoglobin denatures very readily, and it is possible that the discrepancies among some of the physical data are due to some aggregation preceding precipitation of the denatured protein. Measurements on myoglobin solutions reported up to 1943 were summarized by WYMAN and INGALLS[44]. The balance was already at that time in favour of molecular weight 17,000, but a re-examination

of the constants was undertaken when the crystallographic studies of KENDREW[45] began to show interesting results. The same preparation as that examined by Kendrew was used for osmotic pressure measurements. The mean molecular weight calculated from several measurements was 19,000. This value is certainly an upper limit as small amounts of a heavier impurity were always found on ultracentrifugal examination. Slight leakages through the osmometer membrane, which occur very easily with such small protein molecules, would also cause a reduction in pressure, resulting in a too high value for the molecular weight. We therefore take the molecular weight to be 17,000, as the iron analysis will give the most accurate estimate. A careful examination of a 0·6 per cent solution in the ultracentrifuge gave a sedimentation constant of $2{\cdot}2 \times 10^{-13}$ cm/sec. As previously stated, for an accurate evaluation of constants, measurements over a range of concentrations should be carried out. It is justifiable, however, to assume that this value is accurate to $\pm 0{\cdot}1 \times 10^{-13}$. Taking into consideration various previous values for the sedimentation constant summarized by Wyman and Ingalls a mean value of $S_{20}=2{\cdot}1 \times 10^{-13}$ cm/sec is used here for the calculation of the frictional ratio and possible values for the shape and hydration of myoglobin molecules. Assuming the value of 0·74 for the partial specific volume (THEORELL[43]) the frictional ratio f/f_0 calculated from $M=17{,}000$ and $S=2{\cdot}1 \times 10^{-13}$ is equal to 1·05. If the myoglobin molecule is completely spherical this value for the frictional ratio would indicate a hydration of 0·15 gm of water per gram of protein. As this is very near the minimum of a likely value for the hydration, molecules of this protein do not appear to differ greatly from the spherical shape. If the axial ratio of myoglobin were the same as that of haemoglobin, $a/b = 1{\cdot}7$, the hydration would be about 0·08 gm of water per gram of protein. As there is only one free amino group, apart from the ϵ-amino groups of lysine, corresponding possibly only to one polypeptide chain in the myoglobin molecule (PORTER and SANGER[46]) it was thought possible that an asymmetric structure would result. This does not appear to be so, at least in solutions.

Ovalbumin— It is interesting to note that HERZOG[47] obtained a value of $7{\cdot}54 \times 10^{-7}$ sq cm/sec for the diffusion constant of ovalbumin in 1905. This was probably the first measurement to give a good estimate for the molecular weight of a protein. Since that time a large number of physical investigations on solutions of crystalline preparations of ovalbumin were carried out. Though some of the early osmotic pressure and sedimentation equilibrium measurements gave a value of 36,000 for the molecular weight it has been pointed out in a recent note (GUTFREUND[48]) that several recent osmotic pressure investigations

give results which show that the molecular weight of ovalbumin is 45,000–46,000. There is good agreement among the values quoted for the sedimentation constant of this protein[9] $S_{20} = 3{\cdot}5 \times 10^{-13}$ cm/sec. The ADAIRS (G. E. and M. E.)[49] point out that the hydration of ovalbumin is at least 0·32 gm of water per gram of protein and that the molecules are nearly spherical. The frictional ratio, $f/f_0 = 1{\cdot}5$, is the same as that for haemoglobin.

It has already been mentioned in the section dealing with the homogeneity of proteins, that ovalbumin preparations are unstable and become electrophoretically heterogeneous if stored for any length of time. These changes do not appear to affect the homogeneity with respect to size, at least as far as could be observed by methods so far applied to its study. For the three recent osmotic pressure studies reviewed by GUTFREUND[48] ovalbumin prepared by different methods and of different ages were used and gave results in very good accord.

Serum albumin— Crystalline serum albumin from three different sources has been used extensively for physical investigations. In historical order of the investigations these are probably horse, human and bovine. The horse serum albumin investigations were carried out in connection with serological work. Human serum albumin was prepared in large quantities during the war, particularly in U.S.A., as it was found to be a more stable product than whole serum or plasma and was therefore advantageous for transfusion purposes. Bovine serum albumin was also prepared in large quantities during the war, mainly by Armour Laboratories in Chicago, and has since been used for a number of different physical investigations. In fact it has become together with haemoglobin one of the most useful models for the investigation of protein solutions and in the last section of this article the thermodynamic analysis of the osmotic pressure of solutions of these two proteins will be attempted. A large number of different methods of preparation have been used and electrophoretic analysis has shown differences between these. There is however, no evidence that any definite distinction could be made from molecular weight, sedimentation or diffusion data.

ADAIR (G. S.) and ROBINSON[50] and KEKWICK[51] calculated the molecular weight of horse serum albumin from osmotic pressure and sedimentation and diffusion measurements respectively, and both got the value of 70,000 (Kekwick's data were $D_{20} = 6{\cdot}1 \times 10^{-7}$ sq cm/sec and $S_{20} = 4{\cdot}48 \times 10^{-13}$ cm/sec). SCATCHARD, BATCHELDER and BROWN[52] examined human serum albumin by the osmotic pressure method and calculated a molecular weight of 69,000. The same figure was obtained for the molecular weight of bovine serum albumin from the very thorough osmotic pressure measurements of SCATCHARD,

BATCHELDER and BROWN[53]. The diffusion constant is also well established as can be seen from the data quoted in a recent paper by ALBERTY[11], $D_{20} = 6{\cdot}1 \times 10^{-7}$ sq cm/sec, this can also be accepted for the serum albumin from all three sources. A mean sedimentation constant of $4{\cdot}5 \times 10^{-13}$ cm/sec is in fair agreement with all the measurements published so far. It has to be emphasized again that the values for the sedimentation and diffusion constants of various serum albumin preparations are not as accurate as they could be determined with recently refined methods, and it may well be possible that small differences will be detected. No systematic evidence is available to show whether the serum albumins from different animals differ in other respects as the different haemoglobins do.

The shape of serum albumin molecules can be calculated if a value for the hydration is assumed. The ADAIRS (G. E. and M. E.)[54] find a hydration of 0·3 gm of water per gram of protein, from experiments on serum albumin crystals; these are the only data available. If the sedimentation constant is $4{\cdot}5 \times 10^{-13}$, the diffusion constant $6{\cdot}1 \times 10^{-7}$ and the partial specific volume 0·74, then the molecular weight is 69,000 and the frictional ratio is 1·28. Using Oncley's graphs it can be seen that if the hydration is 0·3 gm of water per gram of protein, the axial ratio is 3·5. ONCLEY, SCATCHARD and BROWN[55] give the dimensions of human serum albumin as 150 A/38 A, an axial ratio of 4, but they assume an hydration of only 0·2 gm of water per gram of protein. EDSALL and FOSTER[56] suggest even more asymmetric dimensions, namely molecules 190 A long, but they emphasize that their estimate is only approximate.

Insulin— The study of molecular constants of proteins is of course only interesting as a means to an end. Insulin is a very sparingly soluble protein and is therefore of no value as a model for the study of properties of proteins in solutions. In fact, as it undergoes various remarkable aggregation and disaggregation phenomena, insulin can only be studied by methods which have been well established through the investigation of other well standardized proteins. However these reactions of insulin make it possible to obtain a considerable amount of information about the molecular structure of this protein from studies of its behaviour in solutions.

The first physical investigations on insulin were carried out by SJÖGREN and SVEDBERG[57] in 1931. They studied this protein in the ultracentrifuge, found it to be homogeneous and stable in solutions between *p*H 5–7, and calculated a molecular weight of 35,000 from sedimentation equilibrium measurements. POLSON[58] determined the diffusion constant $D_{20} = 8{\cdot}2 \times 10^{-7}$ sq cm/sec and calculated from this and Svedberg and Sjögren's sedimentation constant,

$S_{20} = 3{\cdot}47 \times 10^{-13}$ cm/sec, the molecular weight $M = 41{,}000$. From crystallographic studies it was quite clear that the molecular weight of insulin must be a multiple of 12,000 (CROWFOOT[59] and recalculated for more accurate values for hydration, GUTFREUND[60]). Up to the publication of a paper by MILLER and ANDERSSON[61] in 1943, it was generally accepted that the molecular weight of insulin is 36,000. Miller and Andersson redetermined the sedimentation and diffusion constants and found the former to be in good agreement with SVEDBERG and SJÖGREN's (Miller and Andersson's $S_{20} = 3{\cdot}53 \times 10^{-13}$ cm/sec corrected for sedimentation in water, while Svedberg and Sjögren's lower value was for sedimentation in buffer). Their diffusion constant, $D_{20} = 7{\cdot}53 \times 10^{-7}$ sq cm/sec, was considerably lower than Polson's and the molecular weight calculated from these data is 46,000. At that time we began a thorough study of insulin in solution, by several methods and under a variety of conditions.

GUTFREUND and OGSTON[35] re-examined the sedimentation constant of insulin and found it to be $3{\cdot}34 \times 10^{-13}$. This was the result of a very careful study, including an investigation of all possible sources of error and is probably accurate to $\pm$ 1 per cent. PEDERSEN[28] remarks that the difference in sedimentation constant between this value and that of former investigators is probably due to differences in the insulin preparation ; this is not likely to be correct as we have examined a large variety of preparations, including some containing about 20 per cent of an impurity of higher molecular weight, and found good agreement among the sedimentation constants for the main component in all of them. The probable causes of errors in previous measurements of the sedimentation constant were pointed out by CECIL and OGSTON[23]. The diffusion constant was also determined by the present author and found to be $7{\cdot}02 \times 10^{-7}$ in solutions of *p*H 7·0 — 7·5 ; from this and $S_{20} = 3{\cdot}34 \times 10^{-13}$, assuming a partial specific volume of 0·75, the molecular weight is 46,700.

From further ultracentrifugal and detailed osmotic pressure studies GUTFREUND[60, 62] found that in neutral solutions of moderate protein concentration (0·5 — 1 per cent) insulin was homogeneous and had a molecular weight of 47,000 —48,000. Dissociation of insulin molecules occurred on dilution to protein concentrations of 0·3 per cent or below, and on change of *p*H to above 8 or below 5. It was shown by the combined effect of dilution, *p*H and temperature changes that insulin dissociates into subunits of molecular weight 12,000 and an approximate value for the energy of dissociation was found (Gutfreund). It seems likely that the subunits of insulin molecules are held together by electrostatic forces. WAUGH[63] made detailed studies of an interesting phenomenon, namely that insulin forms a fibrous precipitate, sometimes

tactoids, if heated in very acid solutions. It is interesting to note that this aggregation is reversible, the precipitate can be redissolved in alkaline solutions and physiologically active insulin recrystallized, which has the same physical constants as before. The results of x-ray investigations on fibrous insulin will show whether this material is built up from repeating units of molecular weight 12,000.

It has been possible to use physical methods for the study of solutions even further for the investigations on the structure of insulin. SANGER[64] has shown how homogeneous polypeptides can be prepared from this protein through the oxidation of the disulphide linkages which combine them within the subunit of molecular weight 12,000. These materials have been used by GUTFREUND and OGSTON[65] for the development of a method for the study of substances of low molecular weight in the ultracentrifuge. Details of this investigation do not fall within the scope of this article.

β-Lactoglobulin— This protein has recently been chosen by CECIL and OGSTON[23] for an investigation into the accuracy of the ultracentrifuge 'since it appears to be one of the best defined and most stable proteins'. The diffusion constant of lactoglobulin was also determined by these workers[66], and they have summarized most data available from studies on solutions of this protein. The standard methods of preparation of lactoglobulin are due to PALMER and SÖRENSEN and are discussed in the two papers by Cecil and Ogston; they find no differences in the products resulting from the two methods.

The first physical studies on lactoglobulin were due to PEDERSEN[67] who found this protein to be homogeneous and calculated the molecular weight of 38,000 from sedimentation equilibrium measurements and a higher value, 41,800, from the sedimentation constant $S_{20}=3{\cdot}12\times10^{-13}$ cm/sec and the diffusion constant $D_{20}=7{\cdot}27\times10^{-7}$ sq cm/sec. Osmotic pressure measurements of GUTFREUND[68] and of JOHNSTON and OGSTON[69] gave results in good agreement; they calculated a molecular weight of 37,800 and 37,300 respectively. It appears now that these estimates for the molecular weight are somewhat high. There are several possible causes for this, some of which were discussed in connection with myoglobin. Recent crystallographic measurements by SENTI and WARNER[70] give a value of 35,500 for the molecular weight of lactoglobulin. BULL[71] calculates a molecular weight of 35,020 from osmotic pressure measurements. From Cecil and Ogston's diffusion constant, $7{\cdot}82\times10^{-7}$, and sedimentation constant, $2{\cdot}83\times10^{-13}$ (both values from extrapolation to infinite dilution) and Pedersen's value for the partial specific volume, 0·75, the molecular weight is 35,500.

From the above data of Cecil and Ogston the frictional ratio

$f/f_o = 1{\cdot}26$ can be calculated. As there are not even approximate data for the hydration of lactoglobulin available, it is difficult to draw any conclusions about its asymmetry. If the hydration is of the order of 0·3 gm of water per gram of protein, the axial ratio would be about 3. It should be possible to calculate the amount of ' bound ' water in the wet lactoglobulin crystal from Senti and Warner's experimental records and PERUTZ's[72] formula.

PHYSICAL CONSTANTS OF BIOCOLLOIDS

The six proteins which were discussed in some detail in the previous section have been chosen as examples to show the agreement and disagreement among results. It has been pointed out that even on these six proteins, which have certainly been given more attention than any other large molecules of biological origin, more accurate investigations are still needed to decide certain doubtful points. Much work on similar lines has been carried out on many enzymes, some of them crystalline, and on immuno proteins. Several of these substances have fairly well defined constants. As mentioned in the introduction, a thorough knowledge of the properties of these biologically active proteins as well as information about their activities and intermolecular forces in solutions would enhance our hopes to get an insight into many processes. In the next section an attempt will be made to give a general picture of investigations which can be carried out on colloidal solutions once the molecular constants of the solutes are known fairly accurately.

In the field of virus particles it is difficult to find a sharp division between molecular entities and more complex organisms. Some of the plant viruses, such as tomato bushy stunt virus and tobacco mosaic virus are nucleoproteins which can be crystallized and behave, under suitable conditions, like the homogeneous proteins described in the previous section. These two viruses have been studied very extensively. As they are very much larger than other proteins (with the exception of some haemocyanins) they show more marked deviations from the ideal laws in solutions and great care must be taken if the methods normally used for smaller proteins are applied to the study of virus particles. As these two viruses are well defined, they have been used as model substances for the study of solutions of such giant molecules. Tomato bushy stunt virus forms spherical particles with a diameter of 290 Å and tobacco mosaic virus forms cylindrical rods 2,800 Å long and 150 Å wide. Their respective molecular weights are 12×10^6 and 40×10^6. These dimensions are derived from crystallographic data, electron microscopy, diffusion and sedimentation constants and light scattering measurements which all give results in good accord.

Tobacco mosaic virus solutions have become particularly interesting as they show some very remarkable structural phenomena ; these will be discussed in the next section. A general review of the size of virus particles was given by MARKHAM, SMITH and LEA[73], though this paper was published in 1942 few of the data contained therein have been revised. The most recent references to physical data on tobacco mosaic virus solutions are given by OSTER, DOTY and ZIMM[74]. The two plant viruses mentioned above are by no means the only ones which have been purified or crystallized, they are only given as well established examples. Much physical work on both animal and plant viruses is in progress. This is, however, not the place to review the properties of viruses but only to mention that they form an important group of large molecules of biological origin.

Most of the fibrous proteins are water insoluble and therefore outside the scope of an article on solutions, there are however, some exceptions such as myosin and other asymmetric molecules which play a role in muscle contraction and blood clotting. These mechanisms are of course arousing a great deal of interest. The major advances in the last few years were on the lines of purification of components and chemical investigations ; these are described by BAILEY[75]. It is very likely that a thorough study of the thermodynamic and kinetic properties of myosin and other muscle proteins in solutions would help in obtaining information about the mechanism in which they take part.

There are several other groups of large molecules of biological origin which can be studied by the methods discussed in this article. Glycogens have been investigated in aqueous solutions and methylated glycogens in organic solvents. From a biological point of view the former group seems more important and references to work on these compounds can be found in a recent paper by BELL, GUTFREUND, CECIL and OGSTON[76]. Though some work on other polysaccharides, protein-carbohydrate complexes and protein-lipoid complexes have been carried out by the same methods, they are only mentioned here as examples for possible further applications. Some nucleotides have recently been investigated by diffusion and ultracentrifugal sedimentation procedures (CECIL and OGSTON[77]). Most of the compounds mentioned in this paragraph are not yet well defined and many are probably changed in dispersion during extraction from biological materials. They are also more or less polydisperse and fairly asymmetric.

The more detailed thermodynamic analyses of solutions of some of the well defined proteins can now be attempted. It will be clear that analogous investigations could be carried out on many of the other materials mentioned.

THERMODYNAMIC ANALYSIS OF OSMOTIC PRESSURE OF REAL SOLUTIONS

Definitions of the terms 'osmotic pressure' and 'ideal solution' and a derivation for the osmotic pressure of ideal solutions were given earlier. For the purpose of the derivations given in this section we shall mainly consider two component systems containing water as solvent and a colloidal electrolyte, protein, as solute ; a third component, which will sometimes need consideration, is a small electrolyte which passes freely through the semipermeable membrane. Unfortunately these definitions and derivations are purely hypothetical and are in need of considerable modification for the case of real solutions. A special case of a real solution is a regular solution which was defined by HILDEBRAND[78] as having an ideal entropy of mixing but also a heat of mixing ; here equation 10 for the partial molar free energy of dilution becomes

$$\Delta\bar{G}_1 = -T\Delta\bar{S}_1 + \Delta\bar{H}_1 \qquad \ldots.(34)$$

Strictly speaking this is of course not possible as the heat of interaction or mixing is bound to cause an increase of the entropy of dilution over and above the ideal value. However, the heat of interaction is often negligible, and regular solutions can behave, within the limits of experimental accuracy, as ideal solutions. For the entropy of solution to be an ideal one the volumes of the individual molecules of one component of a solution must not be greater than twice that of the other (FOWLER and GUGGENHEIM[79]). The solutions to be discussed here have an entropy of dilution, which is greater than the ideal one, as the volume of a solute molecule is several thousand times that of a solvent molecule, and usually a heat of mixing or dilution due to electrical interaction effects. Before an analysis of the entropy and heat of dilution of solutions of colloidal electrolytes is attempted, we shall discuss some other factors which contribute to the deviation of colloidal solutions from the ideal laws and which can easily be eliminated before more complex problems are attacked. These factors involve the partial volume taken up by the solute in solution, the solvation of the solute and the error due to the application of van't Hoff's equation 16 to solutions of finite concentrations.

It is clear from the algebraic transformations in equations 12 to 16 that the osmotic pressure calculated from the latter equation will deviate increasingly from that calculated from equation 13 with increasing solute concentration. If we consider however, that even for the readily soluble proteins, haemoglobin and serum albumin, the upper limit of protein concentration obtainable is less than 0·01 gm mol per litre and the maximum concentration considered in the

present experiments is less than 0·002 gm mol per litre, the error involved in the use of the equation $P = RT.n_2/V$ is negligible.

The correction for the volume taken up by the solute molecules is equivalent to van der Waals's correction b and is overcome in the present work by expressing all solute concentrations as per litre of solvent and not per litre of solution. The solvation correction is analogous to that which is applied to the expression for activity coefficients of strong electrolytes. As some of the water in aqueous solutions is bound to the solute and not available as solvent, the apparent solute concentration is raised. It can easily be shown that for globular proteins, deviations from ideality due to hydration are very small compared with the other effects. A hydration of 0·3 gm per gram of protein will increase the osmotic pressure of a solution containing 100 gm of protein per litre, by about 3 per cent.

Entropy of dilution— The study of thermodynamic properties of solutions of large molecules received its major impetus from workers interested in solutions and gels of rubber and synthetic high polymers. During the last ten years a very large amount of work on the evaluation of the entropy of dilution of solutions of long chain molecules of that type has been carried out. A recent monograph by MILLER[80] describes most of these researches ; other accounts of treatments of solutions have been given by GUGGENHEIM[81], ZIMM[82], FLORY[83], HILDEBRAND[84], SCHULZ[85] and HUGGINS[86].

It has been stated that the ideal free energy of dilution is equal to the ideal negative entropy of dilution and that it can be shown by statistical methods that $\Delta\overline{S_I} = -R \ln N_1$ (equation 11). Though other methods of deriving an equation for the entropy of dilution are more general and rigorous (GUGGENHEIM[81, 87], FOWLER and GUGGENHEIM[79]) a convenient proof for equation 11 is given by the simple application of Boltzmann's probabilities,

$$S = k \ln W \qquad \ldots\ldots(35)$$

where k is Boltzmann's constant and W the thermodynamic probability. Treating a solution as a system with sufficient local order to assign the molecules to points on a lattice (quasi-crystalline), the partial entropy of dilution of the solvent can be derived from the reduction in the number of possible arrangements of the solvent molecules. By definition there is no energy of interaction in an ideal solution, hence all configurations have the same energy and they all have the same statistical weight.

If there are $\mathcal{N}n_1 = A_1$ (where $\mathcal{N}$ is Avogadro's number and n_1 the number of mols of the first component) molecules of solvent and $\mathcal{N}n_2 = A_2$ molecules of solute in the solution, the total number of

molecules, and therefore of lattice points is $A_1 + A_2$. There are $(A_1 + A_2)!$ ways of arranging $(A_1 + A_2)$ molecules on as many lattice points, but as all solvent molecules are equal and all solute molecules are equal the number of possible different configurations and hence the thermodynamic probability of such a system is

$$W = (A_1 + A_2)!/A_1!A_2! \qquad \ldots.(36)$$

From equations 35 and 36 it follows that ΔS the entropy of mixing A_1 molecules of solvent and A_2 molecules of solute is

$$\Delta S = k \ln \{(A_1 + A_2)!/A_1!A_2!\} \qquad \ldots.(37)$$

If A is very large an approximation of Stirling's theorem can be used in the form

$$\ln A! = A \ln A - A \qquad \ldots.(38)$$

to simplify equation 37 and we obtain

$$\Delta S = k\left(A_1 \ln \frac{A_1 + A_2}{A_1} + A_2 \ln \frac{A_1 + A_2}{A_2}\right) \qquad \ldots.(39)$$

Since $k\mathcal{N} = R$ and $n_1 = A_1/\mathcal{N}$ and $n_2 = A_2/\mathcal{N}$

$$-\Delta S = R\left(n_1 \ln \frac{n_1}{n_1 + n_2} + n_2 \ln \frac{n_2}{n_1 + n_2}\right) \qquad \ldots.(40)$$

and the partial entropy of dilution of the solvent

$$\Delta\overline{S}_1 = \partial \Delta S/\partial n_1 = -R \ln \frac{n_1}{n_1 + n_2} = -R \ln \mathcal{N}_1$$

which proves the validity of equation 11 used to derive the relation between osmotic pressure and solvent concentration.

The above derivation of the equation for the entropy of dilution of perfect or ideal solutions serves as a simple demonstration of one of the statistical methods. It does give the correct answer for the ideal case and can be applied to real solutions. Besides the effect of solvation already discussed, the deviations from the ideal solution laws are due to ' non ideal entropy of mixing ' and heat of mixing due to solute interaction. HILDEBRAND[84] and GUGGENHEIM[88] discussed the need for a modified treatment of solutions in which the component molecules are not of equal size and pointed out that in that case solutions would not follow Raoult's law even if the heat of mixing is zero. The difficulties involved in the derivation of the entropy equation for mixtures of components of unequal molar volume are only those of finding suitable models and, at least as long as the interaction energy is zero, not mathematical ones. It now remains to be shown why

solutions of large molecules deviate from the ideal laws, even if intermolecular forces are negligible. Furthermore we shall give some description of the methods by which osmotic coefficients for such solutions can be evaluated and compare the results obtained from calculations of various authors with osmotic pressure measurements on solutions of haemoglobin and serum albumin under such conditions that the energy of interaction is likely to be negligible. Unfortunately the mathematical derivations are too long to be given here, furthermore they could only be useful if they were accompanied by a critical discussion of the various physical models and mathematical techniques employed. Some indications of the approach to the problem will be given below. More detailed descriptions, as well as a number of different expressions for the anomalous entropy of dilution will be found in various specialized articles[80–82, 85]; for a very good general introduction to the problem see BUTLER[96].

The simplest model for macromolecular solutions can be applied to long chain molecules when, in the idealized case, each link of the chain takes up approximately the same volume as a solvent molecule. Going back to the lattice model, used for the derivation of the ideal solution laws, we find that each point on the lattice might be taken up either by a solvent molecule or one link of the chain molecule. For the addition of each chain molecule to the solution many lattice points are occupied and when the number of possible arrangements is being computed it must be taken into consideration that adjacent links in the chain must be taken up by adjacent points on the lattice.

Other idealized models are those of macromolecules of rigid spherical or cylindrical shape. Again each solute molecule will take up many points on the lattice, and those taken up by any one molecule must be in certain relations to each other. Both for solutions of long chain molecules and rigid molecules there will be a considerable amount of intermolecular interference. A large number of lattice points in the space between large molecules will be sheltered, and therefore only available to solvent molecules. This latter effect will become particularly marked in fairly strong solutions where the intermolecular distances approach the molecular dimensions.

The above should show, from a statistical point of view, why the addition of large molecules causes larger entropy changes than the addition of molecules of equal size to the solvent. For comparison with osmotic pressure data on proteins, which are reported below, the evaluation of osmotic coefficients for solutions of rigid molecules by ZIMM[82] and SCHULZ[85] is of particular interest. The osmotic coefficient calculated from their respective equations does agree fairly well with the experimental data referred to by Schulz and Zimm.

Discussion of osmotic pressure data— The osmotic coefficient g is defined as

$$\Delta\bar{G}_1 = -gRT \ln N_1 \qquad \ldots\ldots(41)$$

while the activity coefficient f is defined as

$$\Delta\bar{G}_1 = -RT \ln fN_1 = -RT \ln a_1 \qquad \ldots\ldots(42)$$

where a_1 is the activity of the solvent. The osmotic coefficient is determined by the deviations of the osmotic pressure at finite concentrations from the ideal law. If M is the molecular weight of the solute and C the solute concentration in grams per litre of solvent then the osmotic pressure

$$P = gRT.n_2/V = gRT.C/M \qquad \ldots\ldots(43)$$

The coefficient g tends to unity as C approaches zero ; the molecular weight can be calculated from the extrapolated value for P/C at $C=0$. The experimental data given below were obtained from osmotic pressure measurements on solutions of haemoglobin and bovine albumin. The respective values for the molecular weight of these two proteins are accurately known, 67,000 and 69,000. The slope of the ideal curve for the plot of osmotic pressure against solute concentration can therefore be compared with the experimental data and an attempt will be made to get some information about intermolecular forces in solutions from such results.

It is possible to get an approximate value for the energy of interaction, attractive and repulsive forces between the solute molecules, by subtracting the calculated entropy of dilution from the measured free energy of dilution. This can, of course, not yield strictly accurate results as the energy of interaction would also affect the entropy of dilution. Some attempts have been made to find mathematical solutions for the entropy of mixing components with non-zero energies of mixing (GUGGENHEIM[81], ALFREY and DOTY[89] and MILLER[80]. Experimental approaches will be suggested as a preliminary guide for the differentiation between the effects of interaction energies and non ideal entropy of dilution upon the magnitude of the osmotic coefficient. Sometimes it has been possible to differentiate experimentally between the partial entropy of dilution and the partial heat of dilution by carrying out osmotic pressure or vapour pressure measurements at different temperatures as it follows from Maxwell's relations that $\partial\Delta G/\partial T = \Delta S$. In high polymer systems with a very large entropy of dilution it has been possible to determine the temperature dependence of the partial molar free energy of dilution of the solvent (GEE[90]) but in experiments with proteins this has not been possible so far and it

will be seen from the data presented below that the measured pressure difference from a change of temperature of 20° or so, would be very small.

It has already been shown in Adair's (G. S.) early studies on the osmotic pressure of haemoglobin solutions that under all the many conditions employed the thermodynamic behaviour of this protein in solutions of finite concentration always deviates from the ideal laws. Before Adair's careful investigations the results of osmotic pressure studies showed great fluctuations which were due to an incomplete understanding of the Gibbs-Donnan equilibrium and the resulting partial osmotic pressure due to unequal distribution of small electrolytes. Fortunately, thanks to Adair's achievements, we possess today sufficient knowledge of these Donnan effects to be able to reduce them so as to be negligible or to estimate and allow for them. Protein molecules can of course be strongly positively charged in acid solutions or strongly negatively charged in alkaline solutions. It is thus possible to prepare solutions of colloidal electrolytes of chosen sign and magnitude of electric valence by the use of proteins of well known surface charge, such as haemoglobin and serum albumin, at a given solvent composition. In so-called iso-ionic or iso-electric solutions the hydrogen ion concentration is chosen to give the protein molecules an equal number of positive and negative charges. Though the overall surface charge is zero under such conditions, the total number of charges is at a maximum. The iso-ionic point is the *p*H at which the protein has zero overall charge in pure water, while the iso-electric point is the *p*H at which the protein has zero overall charge in salt solutions, the latter varies with the composition of the solvent. In both haemoglobin and serum albumin the iso-ionic point is at a more alkaline *p*H than the iso-electric points. The ADAIRS (G. S. and M. E.)[91] studied the change of charges on haemoglobin molecules with the addition of ions and discussed the chemical significance of the difference between iso-electric and iso-ionic points. A preferential reaction with certain ions is postulated as the course for these phenomena, both by the ADAIRS and by SCATCHARD[92]. The Donnan effect as well as the interaction between protein molecules and small electrolytes in solutions add to the complications involved in the interpretation of experimental results and great care is required in the planning of experiments in this field. The difficulties can be overcome either by the study of solutions of electrodialyzed salt-free proteins in pure water or in the presence of fairly high concentrations of salt. In solutions of low salt concentrations slight deviations from the iso-electric *p*H as well as salt protein interaction can cause big errors. It should be pointed out that great care has to be taken if experiments in pure water at the

iso-ionic point are carried out as small amounts of contamination with acid or alkaline substances will cause anomalies. If high concentrations of salt are used there may be the danger of some influence on protein dissociation[31]. All these facts give a rather depressing picture of the possibility of getting thermodynamic information about protein solutions from osmotic pressure measurements. These difficulties can of course be overcome easily for molecular weight determinations, by the extrapolation procedure. The number of complications should not deter one however, from this approach to the problem of properties of solutions, but they should impress on us the necessity to collect experimental data under a very large variety of conditions before any unambiguous interpretation is possible.

HARTLEY[93] suggested from theoretical considerations that a 'swamping process' could be used. Quoting from Hartley

> In the case, however, where a swamping excess of a simple electrolyte is used in the osmotic pressure measurement, not only are the simple ions brought to approximate equality in total concentration on either side of the membrane, but it can be shown that they differ from complete equality, in exactly the right amount to compensate for the effect of interionic forces. The colloidal ions now behave, not only as though they were in solution by themselves, but as though in ideal solution, except in so far as non-ionic deviations from ideality, which are important in all solutions of macromolecules, are operative.

An algebraic proof of this statement is adduced by HARTLEY[94]. Experience in our recent experiments has shown that there is an optimum concentration of simple electrolytes at which the deviations from the ideal laws can be accounted for by entropy effects, due to solute/solvent molar volume ratio, alone.

For the purpose of comparing experiments under varied conditions we shall determine the osmotic coefficients of solutions containing 100 gm of protein per litre of solvent. *Figure 1* shows a plot of the osmotic pressure of bovine serum albumin solutions against concentration at 3° C. The dotted line shows the pressure of a hypothetical ideal solution of serum albumin, the full line is drawn through the experimental points obtained. At salt concentrations of 0·2 *M* or above, between *p*H 4·6–5·3, there appears to be no significant difference in the results obtained under the various conditions chosen. It is therefore likely that the deviations from ideality shown by the measurements on albumin solutions, whose records are given in *Figure 1*, are almost entirely due to non-ideal entropy of mixing. (A small correction is to be made for hydration ; this has been neglected in all the graphs shown in this section. In solutions containing 100 gm of protein per litre of solvent the effect of hydration corresponds to 0·03 in the osmotic coefficient.)

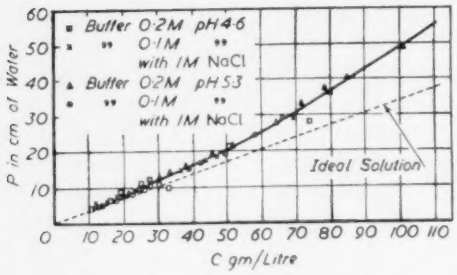


Figure 1. Osmotic pressure of serum albumin

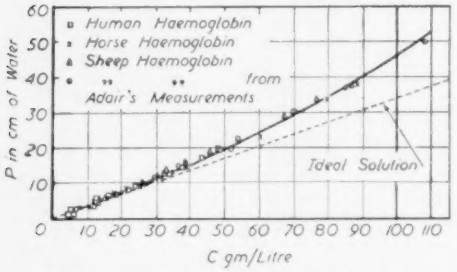


Figure 2. Osmotic pressure of haemoglobin solution in 0·2M phosphate, pH 6·9

Figure 2 shows a plot of the osmotic pressure against protein concentration from measurements on haemoglobin solutions, using the same scale and temperature as *Figure 1*. With this protein it is likely that the conditions chosen, 0·2 *M* phosphate buffer of *p*H 6·9, are the best for avoiding deviation from the ideal solution laws due to ionic causes. Greater salt concentrations appear to cause dissociation of haemoglobin molecules (GUTFREUND[31]).

The osmotic coefficients at 100 gm of protein per litre are about 1·33 and 1·43 for haemoglobin and serum albumin respectively. From SCHULZ's [85] equations the osmotic coefficient for a solution of spherical molecules of the same volume as these two proteins is 1·3. Considering that higher osmotic coefficients are to be expected from asymmetric molecules this agreement is very good as haemoglobin, and even more so, serum albumin, do deviate from spherical shape. It is likely that additional entropy anomalies due to asymmetry of rigid particles only become noticeable at concentrations at which the average intermolecular distance of the colloidal solute has the same order of magnitude as the dimensions of the colloids themselves. It is interesting to consider in this connection, as well as for the analysis of long range intermolecular forces, that in a haemoglobin or serum albumin solution containing 100 gm/litre the centres of protein molecules are, on the average, about 100 Å apart.

It is of course possible that the assumption of conditions of 'ionic ideality' for solutions of colloidal electrolytes in the presence of high concentrations of diffusible ions is erroneous. Some activity effects may influence the deviation of such solutions from ideal behaviour and the good agreement of the observed osmotic pressure with the entropy of dilution calculated by methods of statistical mechanics may be fortuitous. Further detailed theoretical and accurate experimental investigations on the osmotic pressure of concentrated protein solutions

in solvents of varied *p*H and salt concentrations are required. The use of salts of varied activity coefficients at high concentrations might help to throw some light on this point.

Having given a brief outline of the osmotic properties of two model proteins in solutions of high salt concentration, we shall now consider the osmotic activity of serum albumin in pure water. It is well known that most proteins have their maximum solubility in solutions of moderate salt concentrations and if solutions saturated with a protein under such conditions (the optimum salt concentration for protein solubility varies for different proteins) are dialyzed against pure water or strong salt solutions, precipitation or crystallization occurs. Though iso-ionic proteins have zero overall charge, their charges are not uniformly distributed over the surface of the molecules and resulting dipole effects can be observed, to varying extents, from dielectric studies on solutions of different proteins[25]. The effect of this dipolar structure appears to be most marked in solutions in pure water and attractive forces between mutually orientated protein molecules reduce their free energy and can cause formation of crystals. It would be interesting to compare dipolar properties with solubility ; lactoglobulin, for instance, has a very large dipole moment and is almost insoluble under iso-ionic conditions. Work on these lines is at present in progress and it is hoped that the forces between dipolar colloidal electrolytes in water can be derived theoretically and compared with the experimental results for free energy changes. Apart from the few known facts from solubility measurements which are also in need of extension, we have only a few osmotic pressure measurements on solutions of bovine serum albumin in pure water to confirm the attractive forces between dipolar protein molecules. *Figure 3* shows a record of these preliminary osmotic pressure measurements of bovine serum albumin in water (*p*H 5·0–5·1). It can be seen that the osmotic coefficient becomes smaller than unity. For the evaluation of the attractive forces the non-ideal curve (*Figure 1*) must of course be taken as a base line.

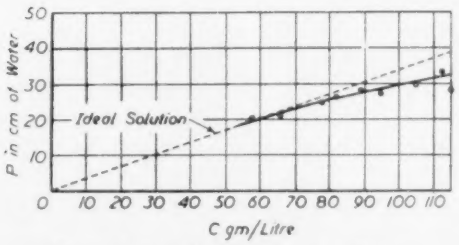


Figure 3. Osmotic pressure of serum albumin solutions in water

A careful study of serum albumin or haemoglobin in very dilute salt solutions is now required to detect any intermediary effects. Unfortunately it is very difficult to get accurate and unambiguous results under such conditions, both on account of ion-protein interaction and due to Donnan effects

arising from shifts of the iso-ionic point to iso-electric points. Of course small net charges on the protein can give rise to big Donnan corrections in dilute salt solutions.

The above discussion on the planning and interpretation of osmotic pressure measurements on solutions of some well defined proteins is given as an indication how this method can be used for the study of forces of interaction between large molecules of biological origin. The field is full of complications, both from the point of view of theoretical prognosis and of interpretation of experimental results. The combination of theoretical and experimental investigations promises however, to give much interesting information on the properties of solutions and on interactions between molecules which play an important part in biological mechanisms. The contents of this section are necessarily only a preface to the description and interpretation of the work now in progress ; as it is written while both experimental and theoretical investigations are in full progress, even many of the data already available have not yet been interpreted and could therefore not be included in this account.

The attention of those interested in long range forces in colloidal systems has so far been mainly given to gels and solutions of tobacco mosaic virus and bentonites. Among protein gels those prepared from gelatine and tobacco mosaic virus have been studied most. As tobacco mosaic virus consists of much more well defined particles and forms gels with crystalline structure a considerable amount of physical measurements and theoretical speculations on this system have been described during the last few years. VERWEY and OVERBEEK[95] review all the current theories on interaction between such colloidal particles but one cannot help feeling that the models chosen for the foundations of the various hypotheses, especially with regard to tobacco mosaic virus, are somewhat arbitrary. The surface charge and charge distribution as well as the change of the partial free energy of the solvent due to the presence of simple ions are susceptible of experimental verification. ONSAGER's[97] theory on the effect of shape on the interaction of colloidal particles, which has appeared in print since the publication of Verwey and Overbeek's book[95], must be referred to, but unfortunately it did not come to hand in time to be discussed here. This field of research will produce many interesting results if theory and experimental investigations on the lines indicated above are combined in a common cause. It is obviously both of great physical interest from the point of view of theories of molecular interaction and the theory of solutions, and gives scope to careful physico-chemical experiments and last but not least, any results obtained should be of great value to those interested in biological phenomena.

REFERENCES

[1] ADAIR, G. S. *Proc. Camb. Phil. Soc. (Biol.)* 1 (1924) 75
[2] OSTER, G. *Chem. Rev.* 43 (1948) 319 ; see also this vol p 73
[3] OGSTON, A. G. *Physiol. Rev.* 27 (1947) 228
[4] KENDALL, J. *Trans. Farad. Soc.* 33 (1937) 1
[5] GUGGENHEIM, E. A. *Modern Thermodynamics* London, 1933
[6] DONNAN, F. G. *J. chem. Soc.* (1939) 707
[7] WIENER, O. *Ann. phys. Chem.* 49 (1893) 139
[8] EINSTEIN, A. *Z. Elektrochem.* 14 (1908) 235
[9] ONCLEY, J. L. *Ann. N.Y. Acad. Sci.* 41 (1941) 121
[10] EDSALL, J. T. *Adv. in Colloid Sci.* 1 (1942) 269
[11] ALBERTY, R. A. *J. Amer. chem. Soc.* 70 (1948) 1675
[12] ANDERSON, E. A. and ALBERTY, R. A. *J. phys. and colloid Chem.* 52 (1948) 1345
[13] OGSTON, A. G. *Nature, Lond.* 157 (1946) 193
[14] MACPHERSON, C. F. C., MOORE, D. H. and LONGSWORTH, L. G. *J. biol. Chem.* 156 (1944) 381
[15] HERDAN, G. *Nature, Lond.* 163 (1949) 139
[16] ALEXANDER, A. E. and JOHNSON, P. *Colloid Science* Oxford, 1949
[17] OAKLEY, H. B. *Trans. Farad. Soc.* 31 (1935) 137
— — *ibid* 33 (1937) 372
[18] JULLANDER, J. *Ark. Kem. Min. Geol.* 21A (1945) No. 8
[19] COULSON, C. A., COX, J. T., OGSTON, A. G. and PHILPOT, J. ST. L. *Proc. roy. Soc.A* 192 (1948) 382
OGSTON, A. G. *ibid.* 196 (1949) 272
[20] KEGELES, G. and GOSTING, L. J. *J. Amer. chem. Soc.* 69 (1947) 2516
[21] LONGSWORTH, L. G. *ibid* 69 (1947) 2510
[22] SVEDBERG, T. and PEDERSEN, K. O. *The Ultracentrifuge* Oxford, 1940
[23] CECIL, R. and OGSTON, A. G. *Biochem. J.* 43 (1948) 592
[24] SCHMIDT, C. L. A. *The Chemistry of Amino Acids and Proteins* London, 1943
[25] COHN, E. J. and EDSALL, J. T. *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions* New York, 1943
[26] TRISTRAM, G. R., PORTER, R. R. and SANGER, F. *et alii Barcroft Memorial Volume on Haemoglobin* London, 1949
[27] JOHNSTON, J. P., LONGUET-HIGGINS, C. H. and OGSTON, A. G. *Trans. Farad. Soc.* 41 (1945) 558
[28] PEDERSEN, K. O. *Ann. Rev. Biochem.* 17 (1948)
[29] KENDREW, J. C. and PERUTZ, M. F. *Proc. roy. Soc.A* 194 (1948) 375
[30] ADAIR, G. S. *ibid* 108 (1925) 627 ; 109 (1925) 292
— *ibid* 120 (1928) 753
— *ibid* 126 (1929) 16
[31] GUTFREUND, H. *Barcroft Memorial Volume on Haemoglobin* London, 1949
[32] SVEDBERG, T. and FÅHRAEUS, R. *J. Amer. chem. Soc.* 48 (1926) 430
[33] COHN, E. J., HENDRY, J. H. and PRENTISS, A. M. *J. biol. chem.* 63 (1925) 721
[34] SVEDBERG, T. and NICHOLS, J. B. *J. Amer. chem. Soc.* 49 (1927) 2920
[35] GUTFREUND, H. and OGSTON, A. G. *Biochem. J.* 40 (1946) 432
[36] BOYES WATSON, J., DAVIDSON, E. and PERUTZ, M. F. *Proc. roy. Soc.A* 191 (1947) 83
[37] WYMAN, J. *Adv. in Protein Chem.* 4 (1948) 407
[38] TISELIUS, A. and GROSS, D. *Kolloid Zeit.* 66 (1934) 12
[39] LAMM, O. and POLSON, A. *Biochem. J.* 30 (1936) 528
[40] ADAIR, G. S. *Barcroft Memorial Volume on Haemoglobin* London, 1949
[41] ROCHE, J., ROCHE, A., ADAIR, G. S. and ADAIR, M. E. *Biochem. J.* 26 (1932) 1811
[42] GRALEN, N. *ibid* 33 (1939) 1907
[43] THEORELL, H. *Biochem. Z.* 268 (1934) 46
[44] WYMAN, J. and INGALLS, E. N. *J. biol. Chem.* 147 (1943) 297
[45] KENDREW, J. C. *Proc. roy. Soc. A.* 201 (1950)
[46] PORTER, R. R. and SANGER, F. *Biochem. J.* 42 (1948) 287
[47] HERZOG, R. O. *Z. Electrochem.* 13 (1905) 533
[48] GUTFREUND, H. *Nature, Lond.* 153 (1944) 406
[49] ADAIR, G. S. and ADAIR, M. E. *Trans. Farad. Soc.* 36 (1940) 23
[50] — and ROBINSON, M. E. *Biochem. J.* 24 (1930) 1864
[51] KEKWICK, R. A. *ibid* 32 (1938) 552
[52] SCATCHARD, G., BATCHELDER, A. C. and BROWN, A. *J. Clin. Invest.* 23 (1944) 458
[53] — — — *J. Amer. chem. Soc.* 68 (1946) 2320

[54] ADAIR, G. S. and ADAIR, M. E. *Proc. roy. Soc.B* 120 (1936) 122
[55] ONCLEY, J. L., SCATCHARD, G. and BROWN, A. *J. Phys. Colloid Chem.* 51 (1947) 184
[56] EDSALL, J. T. and FOSTER, J. F. *J. Amer. chem. Soc.* 70 (1948) 1860
[57] SJÖGREN, B. and SVEDBERG, T. *ibid* 53 (1931) 2657
[58] POLSON, A. *Kolloid Zeit.* 87 (1939) 149
— *Dissertation, University of Stellenbosch* ———, 1937
[59] CROWFOOT, D. *Proc. roy. Soc.A* 164 (1938) 580
[60] GUTFREUND, H. *Biochem. J.* 42 (1948) 156
[61] MILLER, G. L. and ANDERSSON, K. J. J. *J. biol. Chem.* 144 (1942) 459
[62] GUTFREUND, H. *Biochem. J.* 42 (1948) 544
[63] WAUGH, D. J. *J. Amer. chem. Soc.* 70 (1948) 1850
[64] SANGER, F. *Biochem. J.* 44 (1949) 563
[65] GUTFREUND, H. and OGSTON, A. G. *Ibid.* 44 (1949) 163
[66] CECIL, R. and OGSTON, A. G. *Ibid.* 44 (1949) 33
[67] PEDERSEN, K. O. *ibid* 30 (1936) 961
[68] GUTFREUND, H. *Nature, Lond.* 155 (1945) 236
[69] JOHNSTON, J. P. and OGSTON, A. G. *Trans. Farad. Soc.* 42 (1946) 789
[70] SENTI, F. R. and WARNER, R. C. *J. Amer. chem. Soc.* 70 (1948) 3318
[71] BULL, H. B. *ibid* 68 (1946) 742
[72] PERUTZ, M. F. *Trans. Farad. Soc.* 42 B (1946) 187
[73] MARKHAM, R., SMITH, K. M. and LEA, D. *Parasitol.* 34 (1942) 315
[74] OSTER, G., DOTY, P. M. and ZIMM, B. *J. Amer. Chem. Soc.* 69 (1947) 1193
[75] BAILEY, K. *Ann. Rep. Progr. Chem.* 43 (1946) 280
[76] BELL, D. J., GUTFREUND, H., CECIL, R. and OGSTON, A. G. *Biochem. J.* 42 (1948) 405
[77] CECIL, R. and OGSTON, A. G. *J. chem. Soc.* (1948)
[78] HILDEBRAND, J. H. *Solubility* New York, 1936
[79] FOWLER, R. H. and GUGGENHEIM, E. A. *Statistical Thermodynamics* Cambridge, 1939
[80] MILLER, A. R. *Theory of Solutions of High Polymers* Oxford, 1948
[81] GUGGENHEIM, E. A. *Proc. roy. Soc.A* 183 (1944) 203, 213
[82] ZIMM, B. H. *J. chem. Phys.* 14 (1946) 164
[83] FLORY, P. J. *ibid* 13 (1945) 453
[84] HILDEBRAND, J. H. *ibid* 15 (1947) 225
[85] SCHULZ, G. V. *Z. Naturforsch.* 2a, (1947) 28, 348, 411
[86] HUGGINS, M. L. *J. phys. colloid Chem.* 52 (1948) 248
[87] GUGGENHEIM, E. A. *Proc. roy. Soc.A* 135 (1932) 181
[88] — *Trans. Farad. Soc.* 33 (1937) 151
[89] ALFREY, T. and DOTY, P. M. *J. chem. Phys.* 13 (1945) 77
[90] GEE, G. *Quarterly Reviews* 1 (1947) 265
[91] ADAIR, G. S. and ADAIR, M. E. *Biochem. J.* 28 (1934) 1230
[92] SCATCHARD, G. *J. Amer. chem. Soc.* 68 (1946) 2315, 2320, 2610
[93] HARTLEY, G. S. *Quarterly Reviews* 2 (1948) 152
[94] — *Trans. Farad. Soc.* 31 (1935) 106
[95] VERWEY, E. J. W. and OVERBEEK, J. T. G. *Theory of the Stability of Lyophobic Colloids* Amsterdam, 1948
[96] BUTLER, J. A. V. *Chemical Thermodynamics* London, 1946
[97] ONSAGER, L. *Ann N.Y. Acad. Sci.* 51 (1949) 627

2

FUNDAMENTAL STRUCTURES IN BIOLOGICAL SYSTEMS

K. M. Rudall

WHILE here and there we have succeeded in integrating whole groups of phenomena, the properties of living matter still spread out in a vast array not easily comprehended. We know the kind of simplification we require ; it is to be aware of the nature and position of every atom in living structures. This resolves itself into knowing the structure of and the interrelationship between all the molecules. It is well to know that this is our aim and then we can judge, as progress is made here and there, how far along the road we have travelled. In other words, we can see just to what degree we have contributed to achieving the desired goal.

The question often arises in discussion, what are the definable aspects of biophysics ? The physics of large scale phenomena, such as the eye as an optical instrument, auditory mechanisms, pumps, valves and girders, is readily appreciated ; more strange are the numerous electrical phenomena of nerve and muscle. There is also the huge field of study of the effects of ionizing radiations on living tissue. We are concerned here, however, with those physical studies which are capable of elucidating the structure of biological molecules and the physics of small scale phenomena. To emphasize the importance of this subject we can say that knowledge of the structure of such molecules is as important to biology as atomic structure is to chemistry and physics. Physical measurements are bound to reveal the shape of these molecules and how they change their shape, how unique structures are built up from several molecular types and how they behave like minute instruments ; but the fundamental problem in all these conceivable systems of great complexity is the analysis of structure at the molecular level.

There are many views on what constitute the most wonderful things about life. Most important is the nature of the molecules and from this all the others follow. The core of the physical problem is the structure, but not all measurements of properties are an equally reliable guide. We rely mostly on, and desire to perfect, those methods which take us most directly to the structure. Of outstanding importance are the optical methods involving x-ray diffraction, electron microscope image formation, and spectroscopic absorption ; but many other studies are complementary ; they facilitate and expand the results obtained by

the direct methods. The latter, however, must form the dominating approach to the systematic study of structure in biological systems.

In the biological field x-ray diffraction studies have now been applied for nearly thirty years, while electron microscope studies have been usefully applied for only one third as long. While the diffraction studies resolve discontinuities of atomic dimensions, as well as macromolecular dimensions, the electron microscope at present resolves structures only of the order of macromolecules. Thus the two methods overlap in a way that is extremely convenient. The great interest lies in the detail of intramolecular structure ; it is information of a rare kind and of greater philosophic content. Were the resolving power of the electron microscope capable of imaging all structure down to the atom it would be the most important method of all. At present it contributes only indirectly to considerations of intramolecular structure. The diffraction method too has its limitations as it fails progressively as we depart from geometrical perfection.

This article is entitled ' Fundamental structures in biological systems ' in the belief that the subject matter is really basic to all the extraordinary properties of living matter. In considering some of the great contributions which have been made in this field we are concerned mainly with fundamental structures of the following kinds : proteins, polynucleotides, polysaccharides and lipids. Not all of these substances can have special attention here and we are concerned chiefly with the first three classes, which are polymers. The compounds of special interest may be few in number of classes but they are particularly numerous in species. The studies to be described explain certain sources of this variation, giving a proper basis for the classification of various types.

A number of x-ray diffraction photographs are included, representing the principal types of chain molecule which have been recognized in biological systems. The interpretation of these patterns is certainly difficult, but its value is evident and has proved a fruitful guide in many directions. A limited number of references is appended giving access to more detailed discussions.

POLYPEPTIDES AND PROTEINS

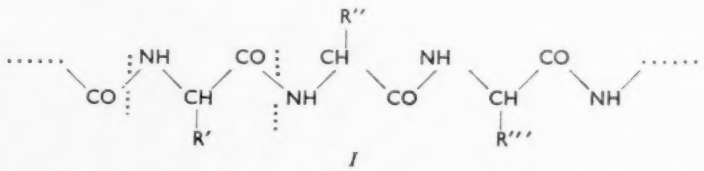
Proteins make life possible ; they also make it bearable or unbearable. Life originated because proteins originated. Where have we arrived so far in the study of these tremendously important molecules ? The field of intramolecular structure was opened up by the study of such unlikely objects as fibres of silk, hair, wool, and tendon. There are still those who do not readily see that the structure of fibres is a very

fundamental aspect of protein studies which is only at the beginning of its profitable advance. So many important things have been discovered that we now know that to pay first attention to protein fibres was a wise step.

Contributions to protein structure have come from a long succession of chemical studies, but from the physical point of view our ideas on intramolecular structure have been created by the comprehensive work of ASTBURY and his colleagues, comprehensive because it ranged widely in the biological field and recognized the main principles of intramolecular architecture in polypeptide systems. It is these contributions that will be discussed here, what is more and what is less certain about them, and how they affect certain classical problems in biological systems.

Structure of extended β-type polypeptide chains— The simplest fibres consist of a parallel arrangement of molecular chains. Natural silk is composed of a protein called fibroin, and MEYER and MARK first interpreted the axial repetition of pattern in x-ray diffraction patterns as twice the distance between successive amino acids in the constituent polypeptide chains. Studying a wide variety of animal and vegetable proteins, Astbury realized that these could almost always be obtained in a form consisting of extended polypeptide chains bound together into regular bundles or crystallites, so named because such ordered regions give diffraction patterns of the type so well known for crystals of a visible size. The extended polypeptide chains are designated as the β-type because in the beginning a most important property was their derivation from α-type proteins which will be considered later[1, 2]. The β-proteins give us the simpler truths about polypeptide chains as structures and as carriers of reactive groups. The fundamental units of structure in β-type proteins are as follows :

The single polypeptide chain is shown below



I

where R′, R″ *etc* represent the side groups of different α-amino acids, and where on hydrolysis the chain breaks at the position of the dotted lines to give free amino acids. The polypeptide chain may be represented diagrammatically as in *Figure 1*a.

ii The polypeptide grid, which is simply a sheet of polypeptide chains bound together by interaction between side chains as in *Figure 1*b.

iii The three dimensional polypeptide crystallite, which consists of a number of grids piled on top of one another. This pile when viewed along the plane of the grids appears as in *Figure 1*c.

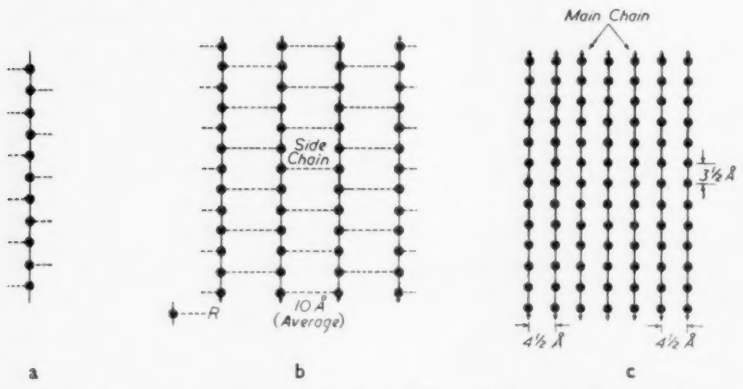
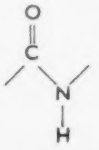


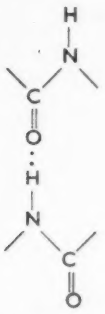
Figure 1. a *Single polypeptide chain* b *Polypeptide grid* c *Polypeptide crystallite*

Many important conclusions follow from the data summarized in these figures. First, quantitative information about the sizes of polypeptide chains is obtained *e.g.* the amino acids follow one another along the chain at distances of 3·4 to 3·5 Å, while the chains are separated laterally in the direction of the side chains by about 10 Å, and laterally between the ' backbones ' of main chains by about 4·5 Å. Furthermore, the above chains are fully extended and cannot be made appreciably longer by stretching. These are all simpler features but a sudden leap into complexities arises from the fact that side groups R′, R″, R‴ *etc* may represent any of some two dozen different amino acids. Only if these residues are arranged in some very regular succession in all parallel chains is there much hope of obtaining information on this order of structure in x-ray diffraction patterns.

The above lay-out of the chain molecules is particularly important in relation to intramolecular forces. In the main chain direction the units of structure are held together by the peptide bond



a strong primary valence bond which is always present. Equally ubiquitous are hydrogen bonds between neighbouring chains



these being in the direction of the backbone linkage in β-proteins. Apart from van der Waals's forces the above C=O · · · H—N bond is the only type of lateral bonding which is common to all extended polypeptides. In polyglycine it would be the only lateral bond, as it is in nylon. Hydrogen bonds of this type are well established by current infra-red absorption studies, but data on their direction is as yet far from extensive. The most likely configuration of neighbouring chains (*Figure 2*) gives the maximum hydrogen bonding between —C=O · · · H—N— groups[2–4]. In *Figure 2* the chains run in opposite directions, thus meeting the requirements of the diffraction data, where there is a pattern repeat of 9·3 Å in the direction of the backbone linkage.

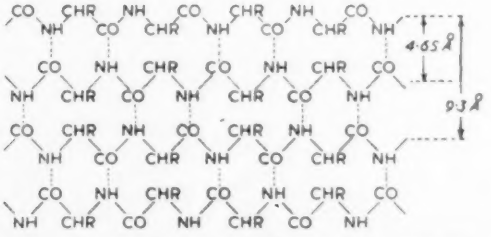


Figure 2. 'Backbone linkage' between extended polypeptide chains

Interaction and bond formation between the side chain groups is undoubtedly the source of the more specific behaviour of protein chain systems. These interactions are three dimensional as is apparent from *Figure 3*.

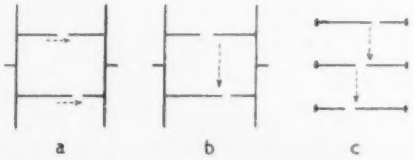


Figure 3. Interactions between side chain groups

For example, there may be interactions between side chain groups of parallel chains in the same grid, as in *Figure 3*a, or interaction between side groups of different 'rungs' of the grid, as in *Figure 3*b. Equally, interactions are possible between side groups of superposed grids as in *Figure 3*c. The possibilities of group interactions are thus especially numerous. At a given rung level one direction of bonding will be favoured when single active groups are present, but which direction this is will depend on the nature of the surrounding groups.

The nature of these bonds will apparently be any of the kinds capable of formation between end groups. The strongest will be of the primary valence type, such as those due to cystine bridges, for the existence of which there is abundant evidence; there are also other conceivable types such as internal peptide bonds. There will frequently be no bonds of these types, at least in the side chain direction of *Figure 3*a, otherwise the spreading of a protein as a monolayer would be inconceivable unless such bonds were broken. A labile bond for which there is much evidence is the electrostatic link between charged basic and acidic groups. Equally we should bear in mind that there could be repulsion of groups of like charge.

Crystallography, density, inextensibility and the chemical analyses all testify to the essential correctness of *Figure 1*a, b and c. The diffraction pattern of stretched keratin, *Figure 4*, is typical of the β-proteins which have complex amino acid compositions. In all these structures there is uncertainty about the general nature of the side chain packing, and we have to accommodate side chains of greatly different lengths from the single hydrogen atom of glycine to the elongated chain of arginine. One can readily calculate that for keratin and myosin long and short side chains could be arranged at the same rung level of the polypeptide grid, leading to a very straight main chain direction. If this is not so and bulky side chains form cross links between main chains, then the main chains must be bent to give a primary fold in the plane of the grid[2]. This situation could be responsible for the reduced axial period of 3·3 to 3·4 Å in keratin and myosin as compared with the 3·5 Å in the simpler structure fibroin, or the theoretical maximum of 3·67 Å which polyglycine seems to approach[5].

Steric hindrance in the arrangement of the side chains will cause departure from a limiting close packed condition. The limiting close packed state should be related to the specific volume of the constituent groups. From the known amino acid composition and using the tables of residue volumes as calculated by COHN and EDSALL[6] the mean residue volume can be estimated. If this is compared with the volume calculated from the x-ray data we find substantial agreement

between the two sets of figures in the case of fibroin and polyglycine, indicating that these two structures are nearly as closely packed as is possible. But estimates for several of the more complex proteins, as in *Table I* show a discrepancy between the mean apparent residue volume and the mean residue volume from x-ray data. This is of the order of 16 Å^3 and is a measure of the degree of non-closeness of packing. There is just enough room for about one water molecule per residue; and sufficient room is available in the proposed residue volume for the amino acids we are dealing with.

Table I. Close Packing of Protein Molecules

Protein	*Mean apparent molal volume of residue, Å*^{3}	*Mean volume of residue from x-rays, Å*^{3}
Keratin ..	133	150
Myosin ..	136	152
Tropomyosin ..	141	155?
Edestin ..	142	
Egg albumin ..	134	150?

From the crystallographic study of β-type proteins we can conceive three subunits : the single polypeptide chain (*Figure 1*a), the polypeptide grid (*Figure 1*b), or a single sheet of chains linked in the backbone direction (*Figure 2*). For only one of these subunits is there any evidence in nature. Monolayers which form at interfaces, such as by spreading of proteins on water, reach a close packed condition where the polypeptide chains are linked in the direction of the backbone and the side chains dip into the water on the one side and stand out from the surface on the other. Such units of structure have often been proposed when considering details of biological structure or as units inside corpuscular proteins. By superposition of actual monolayers, three dimensional arrays have been built up which give ' crystal ' diffraction patterns of the normal β-type. The same concept of a superposed monolayer structure can be derived from the reaction of water with intact β-proteins. Here the water penetrates so as to move the chain-layers apart in the direction of the side chains, a process which, if continued, would lead again to single sheets analogous to monolayers. Free single extended polypeptide chains seem unlikely. Motion of the chain would lead directly to lateral C=O · · · N—H linkage and the single chain would become a small area of a monolayer. A free polypeptide grid, as in *Figure 1*b, is equally unlikely as it could bend upon itself and with C=O · · · N—H linkage give the beginnings of a three dimensional structure. The last two elements, derived from the free polypeptide chain and the free grid, look like the kind of units we might expect inside corpuscular proteins.

Thus it is through the study of chains of the β-polypeptide kind that we begin to understand the principles of protein structure. Individual

chains give every indication of being built on a comparatively simple plan. Outstanding matters on which we hope to learn much more are:

a The kind of repetition of the various side chain groups, R′, R″ *etc.* The little knowledge we do have comes from the analysis of di- and tripeptides, and here in time we may be able to say which acids are adjacent. Related to this is the need for more specific information on the kinds of reactions between side chains.

b We need to know much more concerning the backbone linkage *e.g.* the number and the strength of bonds between —C=O and —N—H groups. Already progress has been made and there is much hope of elucidating details of this sort by infra-red absorption studies[7, 8, 35].

The extended polypeptide or β-type chain is only one of the known configurations among proteins. β-type chains are relegated to a position of no importance by some on the ground that they are denatured and to be thrown away, but structurally they are of first importance, and to say that they give information on half the really important things about proteins may well be a large understatement.

The α-type chain— Whereas it may be said that the chemists envisaged a polypeptide chain of the β-type, the α-type chain is entirely a discovery of x-ray analysis[1, 2, 9], and through these studies its most important properties are recognized, particularly by comparison of its dimensions and behaviour with those of β-type configurations. The classical work here is that on the keratin of hair, wool *etc.* Normal hair gives a large-angle x-ray diffraction pattern with but two principal diffraction spots (*Figure 5*), namely, the usual 9·8 Å reflection at right angles to the fibre axis, and a strong well defined reflection of about 5·1 Å due to a periodicity along the length of the fibres. In themselves these are indeed the scantiest data to go on, but when hair is stretched it gives rise to the β-type pattern, while the 5·1 Å reflection is entirely lost as in *Figure 4*. Under normal conditions the β-pattern reverts to the α-pattern when tension is removed and the fibre allowed to contract to its ordinary length. The α—β transformation is thus a reversible one. The change from $\alpha \rightarrow \beta$ is achieved by stretching the molecules parallel to the main-chain axis, but it can also be brought about by compressing the molecules in a direction at right angles to the fibre axis. Both methods of effecting the transformation are consistent with the extension of a folded configuration of the chains.

These most outstanding properties are characteristic of a number of different proteins. In naturally occurring proteins the existence of a well defined β-configuration is apparently rare. The rather specialized, secreted protein, fibroin, is the classical example, and is a highly insoluble structure. On the other hand the highly soluble protein

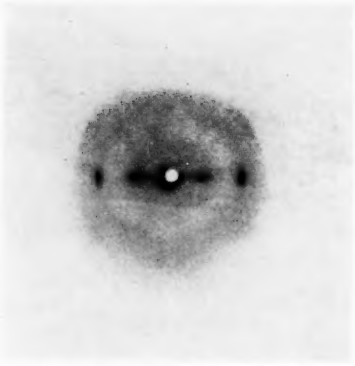


Figure 4. β-protein pattern of human hair stretched by 100 per cent in hot water

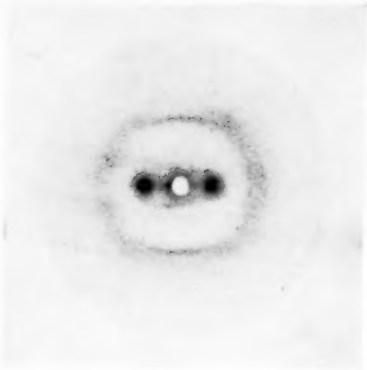
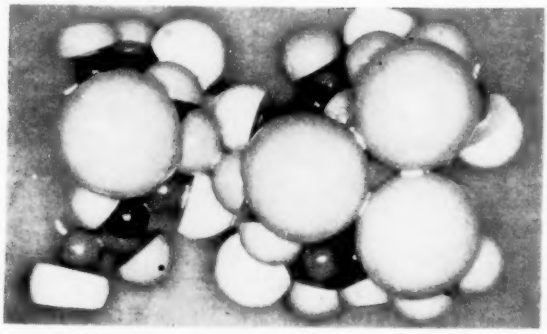
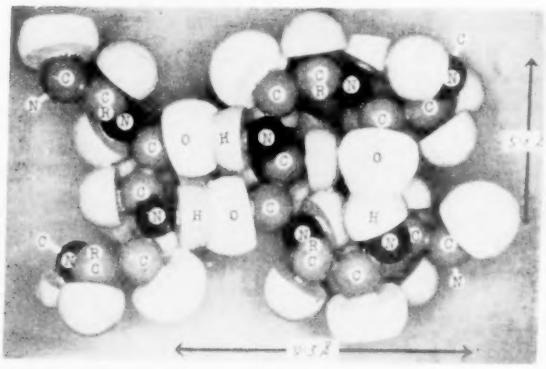


Figure 5. α-pattern of Cotswold wool, not stretched



a



b

Figure 6. Scale models of one complete α-fold and part of another fold of an adjacent chain: a *from one side with the* $-CH_2-$ *groups of the side chains in position,* b *from the opposite side to show the form of the main chain, with side chain* $-CH_2-$ *groups removed*

arthropodin of insect cuticles is another example of a natural β-configuration[10]. Compared with the rarity of natural β-configurations, α-type proteins are very widespread. ASTBURY's classification, *k*, *m*, *e*, *f*, for keratin, myosin, epidermis and fibrinogen indicates the successively discovered sources of typical α-type proteins. Indeed we may reasonably expect them in any body cell, or as a secreted product, and they may yet prove to be the commonest type of polypeptide chain, natural or synthetic.

For investigating the principal phenomena of the $\alpha \rightarrow \beta$ transformation, hair fibres are nearly ideal structures. Their principal advantages are that the asymmetric molecules are naturally highly oriented parallel to the long axis of the hair, and the hair is stabilized in a manner that allows the structure to be fully stretched towards the breaking point and then recontracted to practically the same initial length.

Other α-type fibres *e.g.* myosin and epidermin, show varying amounts of internal slipping so that the initial length of a fibre is not consistently regained on release after stretching.

Because of these characteristics of hair and wool, some quantitative data on the extent of chain folding are obtained[2]. Hair and wool can be stretched in hot water or dilute alkali to just about twice their original length and at this point they commonly break. This gives one possible relationship between the dimensions of the fold (α-type) and the fully extended (β-type) chain *i.e.* the extended chain is approximately twice as long as the α-chain. Another possible relation lies in the fact that the β-form begins to be noticeable at 20–30 per cent extension, and it was originally thought that the degree of unfolding produced an increase of length of about 30 per cent, in agreement with the main periods in α and β patterns of 5·1 Å and 6·8 Å, respectively[1]. On this basis it was impossible to explain the complete extensibility of the fibre by 100 per cent, which was subsequently discovered[2]. To accept the 30 per cent extension during the $\alpha \rightarrow \beta$ transformation in the 'crystalline' phase necessitated the additional hypothesis that other phases of the fibre extended reversibly by amounts equal to or greater than 100 per cent. The simplest view was to regard the process of $\alpha \rightarrow \beta$ transformation as giving the whole 100 per cent extension. The data at our disposal are of a kind that would generally be left unconsidered, but the thrust of Astbury's discussion is that he has believed such data to hold the key to really important aspects of protein structure, and that they hold the key to the detailed mechanism of muscle contraction. Disbelief may result from investigation, and beforehand it may sharpen the aim, but ardent belief in the interest of the subject is to be prized above many other things.

There is difficulty enough in taking the first step, that of proposing

an unfolding from one regular state to another, but there is no other view that accounts for the long-range elasticity, the diffraction data and related phenomena all at the same time. The α and β forms show prominent repetition of pattern along the fibre axis, at 5·1 Å and 6·8 Å respectively. The 5·1 Å distance cannot contain a single amino acid, which repeats at a distance of only 3·4 Å. It may contain 2, 3, 4 *etc* residues and the mean distance along the axis occupied by such residues is given in *Table II*, together with the percentage extension necessary to achieve chain unfolding so that each residue repeats at a distance of 3·4 Å.

Table II. Unfolding of Protein Chains in Hair and Wool

Residues per 5·1 Å period	*Length of residue*	*Extension on unfolding per cent*
2	5·1/2	33
3	5·1/3	100
4	5·1/4	170

Where three amino acids are folded into a distance of 5·1 Å the extension obtained is 100 per cent, and these figures are used as the basis of the proposed α-model. There might be only two or again a larger number of residues per fold. The simplest and most powerful argument against two residues per fold is the certainty of halving planes that would greatly reduce the intensity of the 5·1 Å reflection. In fact, however, the reflection is as a rule very strong indeed. In the β-diagram the equivalent halving planes reduce the strength of 010 to a negligible value, and there is no evidence whatever that alternate side chains are of enhanced scattering power. It might conceivably be argued that in the α-form every other side group was heavier and that on transformation to the β-form the chains slipped over one another so that on the average the weight of the side chains became the same at all levels. We should then have to suppose that this slipping was exactly reversed in the return to the α-form during contraction. These considerations speak powerfully against the model with two residues in each repeat of 5·1 Å, and thus all the more support is given to the model with three residues per fold.

Simplified diagrammatic representations of α and β chains are given in *Figure 7*, while photographs of the actual models of the single α-fold are given in *Figure 6*, the two pictures a and b being taken from opposite sides of the fold. The models were constructed according to known interatomic dimensions, some being covalent radii and some van der Waals's radii, using bond angles which are approximately tetrahedral[9]. Thus this fold is a possible one because interatomic distances and bond angles are satisfied and there is sufficient room for the packing of the side chains. It is also important, however, to have reason for believing it is a stable structure. The basis for the stability is that the loop of the fold is formed approximately at right angles to the direction

in which the side chains project, and is kept in this position in two ways. First and most important, the CH_2 groups of the side chains fit together in close packing. Secondly, the position of the main chain

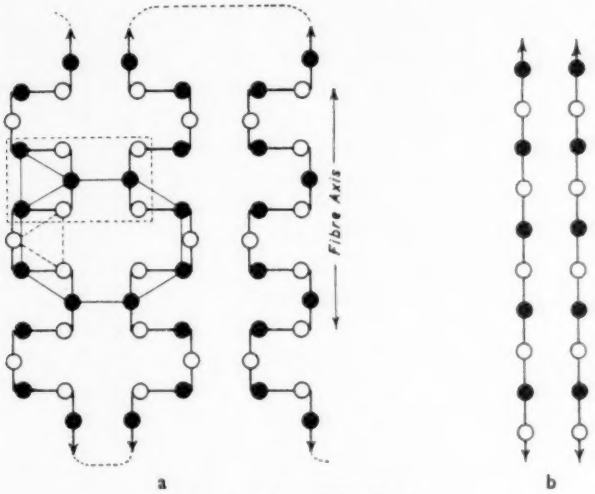


Figure 7 **a** *α-chains and* **b** *β-chains*

loop could be maintained by cross linking of the side chains of neighbouring main chains in the grid. Furthermore, hydrogen bonds, as illustrated in *Figure 6*b, are possible within the folds ; recent studies of porcupine quill tip by infra-red absorption indicate the presence of such bonds directed approximately along the fibre direction, as required by the model[8, 36]. Thus there is a variety of factors associated with principles of close packing, cross linkage and internal bonding which combine to make such a fold stable.

The kind of fold illustrated in *Figures 6* and *7* is the best working hypothesis we have. HUGGINS[11] has proposed a similar three-residue fold model without giving adequate reasons why it should be stable in the wide variety of cases in which it is known to occur. In the Astbury model the remarkable constancy of the 001 spacing during the $\alpha \rightarrow \beta$ transformation is accounted for by having the fold at right angles to the direction of the side chains.

The whole interpretation of the α-configuration depends in the first place on its relation to fully extended polypeptides of the β-type, but even were the latter unknown it seems almost possible to arrive at an independent interpretation of the α-fold on the basis of scale models, principles of packing and intramolecular bonding. From the point of view of the theory of long range elasticity, comparison of the α and β

forms is of the most absorbing interest. There are certain major differences between the two with respect to the intramolecular bonding. First, as regards hydrogen bonds, the form of the β-chain, as illustrated in *Figure 2*, shows the possibility of extensive uniform linkage of —C=O · · · HN— groups, the backbone linkage. The situation is very different in the α-chain. Here a wider variety of hydrogen bonds are possible, principally those within the folds of individual chains and those between chains. Secondly, major changes in the bonding are likely, due to the greatly altered arrangement of the side chain groups. In *Figure 8*, certain side groups AA or BB on the same side of neighbouring β-chains have moved a considerable distance away from one another, while all the side groups within the same chain have moved closer together, from 6·8 Å to about 5·1 Å apart. ASTBURY[9] has attempted to find the cause of long-range elasticity in this situation, supposing that most of the polar groups lie on one side of the chain and most of the non-polar groups on the other side, and indeed in the proteins keratin and myosin the number of polar and non-polar groups are reasonably equivalent. There is criticism of this attractive suggestion on the grounds that in wool both non-polar pairs and polar amino acid pairs do occur as dipeptides[12], but the idea is not disposed of and it represents the search for some unique arrangement of amino acids that could make the α-form more probable than the β-form. Some unique arrangement is suspected, for of the many proteins which can exist in the β-form, only some can exist in a well defined α-form.

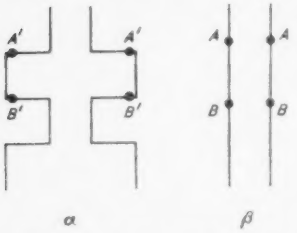


Figure 8. Position of side chains in α- and β-chains

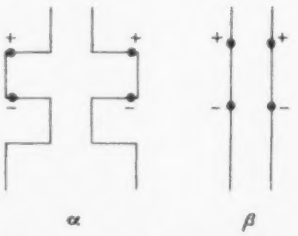


Figure 9. Effect of charged groups on folding of α- and β-chains

Other factors can be thought of which might contribute to the stability of the fold on the one hand, or on the other hand lend force to the contractility of the β-form. If there are oppositely charged side groups at positions as in *Figure 9* these will hold the fold against unfolding tendencies, but when stretched to the β-form the apposition of like-charged groups will oppose the stability of the β-system, which nevertheless could become stable by slipping of grids over one another.

Supercontraction phenomena— The α and β configurations of the same polypeptide chain give us two principal forms in which the distributions of active groups differ. We also want to know what other possible

configurations there may be in order to extend fully our awareness of the range of interchangeable forms. There is a large amount of data on the phenomenon of supercontraction which is recognized since the discovery that wool and hair under certain conditions contract considerably below their normal equilibrium length[2]. This phenomenon excited all the more interest since it bears such a striking resemblance to aspects of muscle contraction[13, 14]. The possibility of a contractile mechanism involving a change of naturally occurring β-polypeptides to the α-configuration has long been considered, but there is no direct evidence for this mechanism, and here we may concern ourselves with what experimental evidence there is for special chain configurations representing a highly contracted condition shorter than that of the α-form.

It has been stressed already that hair and wool form rather ideal structures for the quantitative study of elasticity phenomena. These keratin structures possess a recognizable normal equilibrium length in which the α-type configuration is well defined. A remarkable reversible contraction was discovered by WHEWELL and WOODS[15] as a result of treating hair and wool in cuprammonium solutions. The fibres contracted by about 30 per cent but regained their initial length on washing out the copper with acid. Very striking was the fact that the α-pattern disappeared, but this was not replaced by any other regular pattern which could be related to a superfolded state of the chains. The cuprammonium solution had the general effect of dispersing the constituent chains of either α-keratin, β-keratin or feather keratin and giving rise to an amorphous type of pattern independent of the state of contraction. The mechanism of shortening is of great interest, but we are unable to deduce anything from the comparison of the diffraction patterns of the contracted and uncontracted states. It is probable that the regular α-pattern disappeared, not because component chains underwent further folding, but because these chains were forced out of the normal regular arrangement as in excessive swelling.

A different type of result has been obtained in the thermal contraction of labile α-type proteins[16]. Fibres made from myosin, fibrin, and epidermin, the principal fibrous proteins within mammalian epidermal cells, undergo contraction in hot water at temperatures above 40° C for myosin and fibrin, and above 50° C for epidermin. As a result, the normal α-configuration is replaced by a type of β-pattern. This differs from the normal β-pattern obtained by stretching keratin or myosin, where the extended polypeptide chains lie parallel to the direction of stretching, which we refer to as parallel β, or for short $//\beta$. In thermal contraction the structure is such that the polypeptide chains lie at right angles to, or across, the axis of the fibre, and the x-ray

diagram is thus designated as the cross β pattern or simply Xβ. The three distinctive types of diffraction pattern, α, //β and Xβ are associated respectively with the normal, the stretched and the supercontracted states. We are concerned here with the nature of the Xβ form and whether it gives any clue to the possible superfolding of polypeptide chains.

In a system such as the epidermis, the isolated epidermal protein, or myosin, the cross β pattern is produced in the following manner. Previously oriented material is contracted by placing wetted-out specimens in hot water of temperature 90 to 100° C for a minute or two. Two things occur during the treatment: molecular orientation is completely lost and the intramolecular pattern is changed from the familiar α-form to a disoriented β-type structure. If the material is now stretched in cold water to achieve parallel orientation of internal asymmetric units, the cross β form of diffraction pattern is obtained, *Figure 10*a. In epidermal protein an extension of about 80 per cent was found to give the best results. At extensions above 100 per cent the cross β pattern began to be changed into a mixture of //β and Xβ. The most important characteristic of the cross β pattern is that individual chains are fully extended at least over distances of several residues and that the same kind of backbone linkages occur. In *Figure 10*a, it is evident that the side chain reflection 001 is much weakened, and it is likely to be the same kind of effect as previously described[2] in the setting of //β patterns where there is disturbance of the regularity of structure in the side chain direction with spreading along the layer lines from 111.

The most interesting explanation of the cross β pattern is that which supposes a folding of the main chains upon themselves. The derivation of Xβ from α is given very diagrammatically in *Figure 14*. In b the lengths C—C are occupied by fully extended polypeptide chains, while the repeat of pattern C—D now becomes the familiar backbone period of 4·65 Å. It is highly probable from a consideration of the factors making for stability of the α-fold that some folds are more stable than others. Thus the transformation of α to Xβ could proceed on the supposition that certain folds as at x resist unfolding while thermal agitation readily unfolds the regions y. Regions C—C would then be bonded by the usual backbone linkage and would be permanently or temporarily set according to the sum total of linkages between backbones and side chains.

By postulating that sometimes only a proportion of the chains in a structure adopt the configuration of *Figure 14*b any contraction up to 70 per cent could be obtained on the basis of every third fold remaining intact. In the experimental material there is always loss of orientation

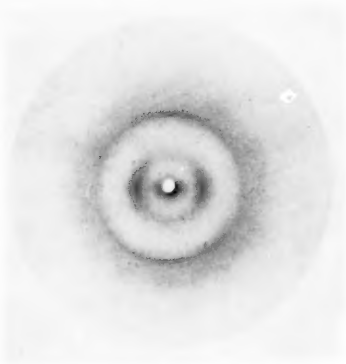


Figure 10a. Cross β-pattern of epidermal protein supercontracted in hot water and stretched 80 per cent

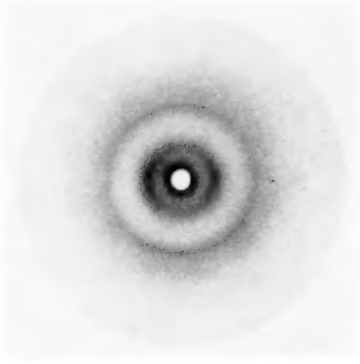


Figure 10b. Cross β-pattern of epidermal protein contracted iso-metrically at 85°

Figure 10c. Pattern of Figure 10a reconverted to the α-form by means of saturated urea

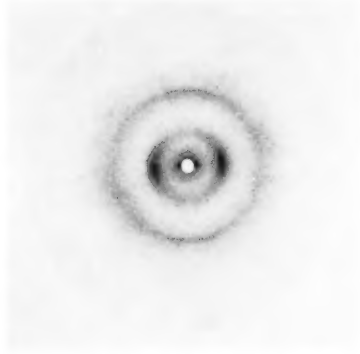


Figure 11. Pattern of the cross β-type produced in insulin as a result of denaturation in thioglycollic acid and urea

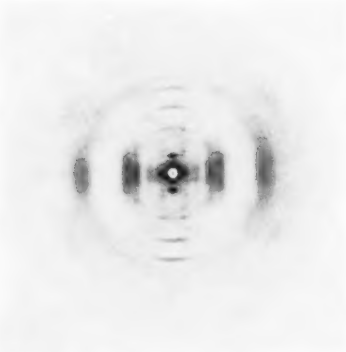


Figure 12. The feather-keratin pattern of seagull quill

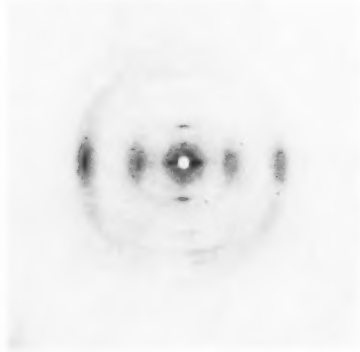


Figure 13. The keratin of the claw of the reptile Varanus niloticus

during isotonic thermal contraction, hence no measure of the true degree of chain shortening is obtainable. But where thermal contaction is carried out isometrically, the cross β pattern appears with its characteristic orientation with reference to the fibre axis as in *Figure 10*b. This corresponds to a direct transformation from α to $X\beta$ with a parallel development of tension. From the beginning this change is the sudden appearance of $X\beta$ replacing α. In the presence of strong urea solutions the cross β form readily changes over into the normal α-form, *Figure 10*c. There seems to be a contracted condition that is reversible, and the importance of this is obvious if we are to be able to carry over ideas about folded chain systems to the reversibility of contraction as it occurs in muscle.

For the present we can summarize the transformations between the various forms of chains as follows. The normal transformation by stretching has the following cycle $\alpha \rightarrow // \beta \rightarrow \alpha$. The full range of the transformations has the form $\alpha \rightarrow X\beta \rightarrow \alpha \rightarrow // \beta \rightarrow \alpha$. The section of this cycle which we should like to resolve is $X\beta \rightarrow \alpha \rightarrow // \beta$ and especially to know whether a simple stretching process will take us from $X\beta$ to α. The usual result of stretching the $X\beta$ form is to produce an increasing amount of $// \beta$, and this is almost inevitable if we choose the model of *Figure 14*b. It is a question of finding an ideal material which when placed in an appropriate medium will show a change from $X\beta$ to α with a real increase of fibre length.

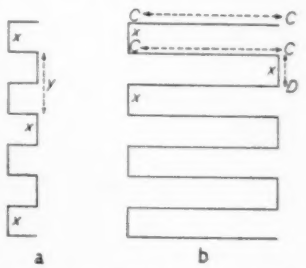


Figure 14. Derivation of Xβ from α-chains

In a very wide variety of proteins the cross β pattern is obtained as a result of thermal treatment, and the common result of stretching this form is to give rise to the $// \beta$ pattern. The first dramatic case was egg albumin which in the poached egg state gives the cross β form[17]; or starting from crystalline egg albumin, denatured in detergents but not heated, the fibres which are only slightly stretched are of the $X\beta$ type, while those highly stretched in live steam give excellently defined diffraction patterns of the $// \beta$ type. That the albumin loosened by the detergent is in the $X\beta$ form is possibly related to a folded condition of polypeptides in corpuscular molecules. High stretching in live steam (*i.e.* without excessive swelling) completes the unfolding of residual folds. In insulin denatured in thioglycollic acid and urea the structure again has the same $X\beta$ form, *Figure 11*. Thus portions of corpuscular protein molecules when denatured by a variety of means *e.g.* heat, detergents, —S—S— reducing agents plus urea, give a common

polypeptide structure of the cross β form. For the present it seems reasonable to relate this to the kind of superfolding proposed as a model for the supercontraction of α-type proteins[16].

Corpuscular proteins— The more recent history of concepts about fibrous and corpuscular proteins may be summed up as follows. Some regarded protein structure as one whole subject with fibrous and corpuscular proteins of equal interest but placing emphasis on the fundamental importance of chain molecules when it comes to intimate detail. Others regarded the fibrous state as of negligible interest as far as corpuscular proteins were concerned. This division seems to have its origin in the vastly different properties of the two types of structure. The well known fibrous proteins had mainly mechanical functions and were considered dull. The corpuscular proteins had the most exciting properties, as in the enzymes which perform almost every conceivable function. These differences in properties naturally indicated differences in structure, but did not necessitate a plan of structure for corpuscular proteins that was unrelated to that in fibrous proteins.

That there is continuity of intramolecular plan in fibrous and corpuscular proteins is indicated by the following :

1 There is no significant difference between the densities of fibrous and corpuscular proteins.

2 On denaturation by heat or other methods, fibrous and corpuscular proteins transform to the same system of polypeptide chains of the β-type.

3 The corpuscular egg albumin molecule behaves like a pile of four monolayers of a fibrous protein[18]. The structure of these monolayers spread as surface films shows them to be of the β-type[19]. Within the molecule the component chains could be in the α-configuration.

These statements amount to the conclusion that the intramolecular arrangements in corpuscular proteins are of the same general kind as those found in fibrous proteins.

Many of the reversible transformations between corpuscular and fibrous conditions will involve little intramolecular change. Examples of this are fibrous and corpuscular insulin, fibrous and corpuscular actin[20], and the reversible change from crystalline to fibrous tropomyosin[21]. In the last example the fibrous state gives the familiar α-type pattern, and it is very probable that this intramolecular arrangement is still present in the crystallized corpuscular form. For more direct information concerning the kind of intramolecular arrangement in a corpuscular protein we have the dimensions deduced by PERUTZ for methaemoglobin[22]. Here the component rods have an axial period of

ca 5 Å, and side dimensions of 10 Å. This is practically identical with the form of the single α-protein chain.

X-ray studies on fibrous and corpuscular proteins— The x-ray studies of fibrous proteins have informed us most directly about intramolecular arrangements and dimensions. By contrast the study of crystalline proteins has mainly provided another means of measuring the molecular weight with the addition of information about the symmetry and packing of the molecules. All the intramolecular detail is also observable but so far it cannot be interpreted except by comparison with the principles of structure revealed by studies of fibrous proteins. In the diffraction patterns of many fibrous proteins longer periodicities are common, and these overlap with the intermolecular dimensions appearing in the diffraction patterns of corpuscular proteins. The electron microscope has given another view of the very same macroperiods in protein fibres, notably of course at the highest end of the series of periodicities, which in their full range probably extend down to atomic dimensions. Reducing all these observations to their simplest form it would seem that many naturally occurring protein fibres are composed of regular subunits which are indistinguishable from corpuscular proteins or are derived from them[23].

Structures of the feather keratin type— The chief protein of birds' feathers is the prototype of a distinct sub-branch within the family of proteins. Feather keratin, as it is called, bears a close relationship to the typical β-protein type but has an axial period of only 3·1 Å as compared with about 3·4 Å. Feather keratin is unique in that it can be stretched reversibly, giving in the stretched state an axial period nearly as great as that in fully extended β-proteins[24]. The mechanism of this limited reversible elasticity is unknown. Feather keratin is also notable in that it generally shows along the fibre axis a remarkable perfection of structure from periodicities of about 94 Å down to atomic dimensions. Similar perfection is also shown in directions at right angles to the fibre axis. These features indicate a highly pure protein species approaching the perfection of detail so characteristic of crystalline proteins. There are, however, many proteins in which this same type of chain structure is present while most of the detail is lost due to intrinsic distortions or admixture with different components. The configuration of the typical feather keratin molecule is likely to represent one of the fundamental structures. Its position with reference to other types is reviewed below.

So far the feather keratin type of molecule has been found only in reptiles and birds[25]. Thus at the molecular level we obtain a further illustration of their affinity, for the characteristic structure of the hard keratins is the same in the two classes. There are, however, specific differences between details of molecular structure in reptilian and avian

keratin and thus we have a well defined case of a plan of structure common to a large group of animals but with specific changes in the subsections of the group. Some of the specific differences within the same kind of cytoplasmic protein can be seen and measured. The keratin of seagull quill, *Figure 12*, and of the claw of *Varanus niloticus*, *Figure 13*, are illustrated. The differences lie in the kind of periodicity within the molecule and probably have their origin in varying arrangement of amino acids.

Among vertebrate epidermal structures the feather keratin type is restricted to the harder tissues of reptiles and birds *i.e.* claw, beak, scale and feather[25]. By restriction can be meant at present only that it is greatly predominating in these structures, but it cannot be said not to occur in traces in other epidermal structures. Particularly remarkable is the manner in which the feather keratin molecule appears to replace the much more frequently occurring α-type structure. Two kinds of replacement occur. First, in the Saurian scale, *Figure 15*a, the outer compact layers have the feather keratin structure and the lower loose layers the α-keratin structure. It would seem from this that there is an alternation of cell generations, one generation containing α-keratin and the other feather keratin. An alternative to this view is that the generative cells forming the scale divide into two layers of different genetic constitution, while by secondary division the compact and loose layers are formed. This latter view corresponds to the situation occurring in the production of a feather from the collar cells, *Figure 15*b. The cells arising from the stratum intermedium synthesize the feather keratin structure, while the stratum cylindricum and the cells destined to form the outer layers of the feather sheath, the

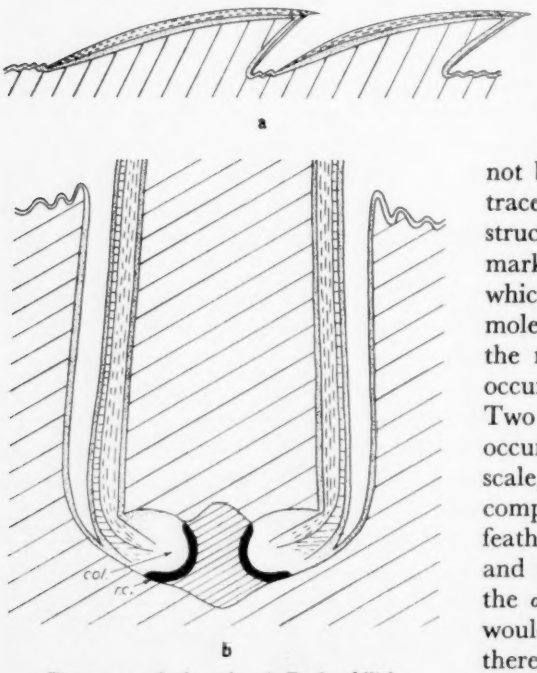


Figure 15. a *Snake scales.* b *Feather follicle*
Feather keratin dashed; α-keratin stippled

calamus sheath and the follicle wall sheath, produce predominantly an α-keratin structure. These remarkable phenomena suggest that different factors for cytoplasmic protein synthesis are segregated in the stratum intermedium cells on the one hand and in the outermost cells of the collar region (col) on the other, the factors being present together in the regeneration cells (r.c.). What are the differences between the postulated factors, and are the same kinds and numbers of amino acids used in each? The possibility that feather and α-keratin are transformations or mechanical isomers of the same polypeptide chain system has been examined. All attempts to transform feather keratin into α-keratin by —S—S— reduction and urea relaxation have been unsuccessful. Nevertheless, the possibility of such a transformation remains. By contrast, it is comparatively easy to convert feather keratin into a typical fully extended β-system.

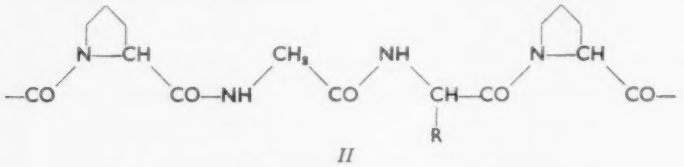
In certain epidermal structures of birds and reptiles the diffraction patterns are consistent with an intracellular mixing of α- and feather keratin type cytoplasmic protein. In the best defined α-type patterns, as from the elastic epidermis between the toes of birds, some meridional periodicities are present which belong to the feather keratin system. It is possible that in this material the extracted cytoplasmic protein may be divisible into a large α-type portion and a smaller proportion of the feather keratin type. This association would bear a resemblance to that of myosin and actin in muscle tissue.

Actin— This important constituent of muscle tissue, first isolated and characterized by STRAUB, has been classified with the feather keratin series on the basis of its diffraction pattern (*Figure 16*). The structure bears no relation to α-type proteins, but rather has the general features of β-type proteins[26]. Furthermore, it is richly provided with the series of meridional periodicities which are so characteristic of the feather keratin pattern. In actin the observed meridional series ranges from 3·6 to 27 Å, which is in simple proportion to the range 3·1 to 23·5 Å of feather keratin. Hence actin could belong to the feather keratin series, in which a number of forms occur with different axial periods, merely as a result of stretching. The type of molecule in the *Varanus* claw (reptilian) is a stable structure with a period of 98·5 Å, which seems to be intermediate between that of true feather keratin and actin.

In the mixed system of the dried muscle fibre the patterns of actin and myosin are superimposed. The long-spacing series was originally thought to represent macroperiods within the myosin molecule, but the small-angle diffraction from intact muscle is apparently identical with that of actin and is absent from the patterns of isolated myosin which contain little actin[27]. In turn these conclusions give point to the following question. Are the closely similar small-angle periods in other

α-type structures *e.g.* porcupine quill and various wool fibres, due to a separate actin-like component rather than to a system intrinsic in the α-component? The whole interpretation of these things is called into question as a result of having to interpret the total pattern of muscle in terms of two distinct protein species.

Collagen type structures—The structure of the great majority of proteins is such that they transform readily into a system of fully extended chains of the β-polypeptide kind. A few types seem unable to reach this end state so readily and some never even approach it. This last distinct group includes all proteins of the collagen class. This type of molecule has two outstanding features. It shows no intramolecular transformation as a result of stretching, yet the main axial period is short *i.e.* 2·8 to 2·9 Å. The first serious attempt to find the underlying principle of structure is due to ASTBURY[28]. He took the above facts, together with the knowledge that collagen and gelatin are particularly rich in proline residues, and arrived at the chain molecule *II*:



II

The success of this model is that it explains the apparent inextensibility, and the short repeat of pattern of 2·86 Å compared with the 3·4–3·5 Å so commonly met in other proteins. One basis on which the chain *II* could be immediately disposed of as a reasonable suggestion would be a wide departure from the required amino acid sequence

—P—G—R—P—G—R—P—G—R—

where P represents proline or hydroxyproline, G glycine and R other amino acid residues.

Even if the analysis showed a considerably lower quantity of the crucial imino acids the proposition would be far from disproved, as only an essential part of a complex structure need be on the above form. In the structural formula above the total axial repeat of pattern is 8·6 Å, which is exactly the value given by the first layer line in the wide-angle diffraction pattern, *Figure 18*a. The 030 group is, however, very prominent, so it seems probable that the chains are shifted parallel to one another by a distance equal to one third of this period, and the reason for such a shift may lie in the possibilities of C=O · · · N—H linking between the backbones of the chains. It is an extraordinary fact that collagen type molecules have never been transformed to the typical β-polypeptide system, either by heat treatment or by various

chemical treatments. This is considerable justification for the view that the chain incorporates frequent imino residues. The fibrous protein elastin which occurs in connective tissue along with collagen also contains a considerable proportion of prolines though less than the quantities found in collagen and gelatin. Elastin does not reveal the characteristic collagen diffraction pattern[28], but on the other hand it does not pack neatly as a system of β-polypeptide chains, which in all probability is due to the rather frequent occurrence of imino residues.

Other structures have been proposed, for example by HUGGINS[11], who also allows for the frequent occurrence of proline residues. In this model there is excellent justification for the intensity of the 2·86 Å repeat, but none for the 8·6 Å repeat except by supposing conveniently arranged residues to be of exceptional scattering power. However, there is an interesting feature since, because of the spiral configuration of the chain, the side chain separation is increased to about 11 Å. The side chains are, on the average, considerably smaller in collagen and gelatin than in most other proteins, and for the straight chain model *II* the diffraction pattern should record a similarly small interchain distance in the side chain direction. That the side chain spacing is rather large even in well dried collagen lends support at least to the spiral aspect of the Huggins model.

Native collagen, as in tendons, shows a well developed series of macroperiods based on a unit of about 650 Å which is found by both x-ray and electromicroscope studies. *Figure 17* illustrates the macroperiods in beef tendon fibrils etched with collagenase. There is also periodic structure across the diameter of the horizontal fibril apparently due to a series of particles in pseudo-hexagonal array. This macroperiod and the amino acid repeat of 2·86 Å behave without coordination. Stretching of fibres alters the macroperiod without changing the wide-angle diffraction. Also, changes of humidity bring about corresponding changes in the macroperiods but do not affect the smaller fibre axis period. The smaller period persists strongly, even when the macroperiods are apparently absent. These things indicate that the wide-angle diffraction pattern is given by only part of a larger unit which, when intact, gives rise to the small-angle diffraction pattern.

The distribution of the collagen type molecule in biological systems may be grouped as follows :

1 white connective tissue fibres
tendons of vertebrates
cartilage ground substance
elastoidin (scales and fin rays of fishes)
ithyocol of swim bladders
mesogloea substance of jellyfish.

2 filaments ejected by sea cucumber
egg capsules of the skate
shell membrane of birds' eggs.

3 cuticle of nematodes *e.g. Ascaris*
cuticle of annelids *e.g.* earthworms.

The substances of the first group are related in that they all seem to be formed by mesodermal cells of the general type of fibroblast. The second group appear to be secreted by mesodermal epithelia, while the third group are laid down in an orderly fashion by secretion from ectodermal epithelia. There is no material where collagen forms the major intracellular or cytoplasmic protein, nor indeed is there any evidence that the molecule exists at all within cells. It may be a transformation product occurring outside cells.

The principal function of structures composed of the collagen-type molecule is to give mechanical support or transmit tension as in tendons. In animal tissues the chief substance performing the same functions is chitin, and chitin and collagen appear to replace each other in a peculiar way. The principal fibrous product of the ectodermal epithelium in annelids *i.e.* the cuticle, is a collagen-type structure, *Figure 18*b. By contrast the cuticle of arthropods is composed of the polysaccharide, chitin, and a β-type protein. In this respect the chitin system replaces collagen ; and a further example is the chitinous tendon in arthropods, the analogous structure in vertebrates being, of course, collagen. In annelids only certain local regions produce chitin, as in the bristles and the lining of the crop, while most epidermal cells produce collagen. In arthropods the capacity for producing chitin has spread to substantially all epidermal cells with loss of the collagen-producing function. As far as the nature of the materials used is concerned, this is one of the most striking types of replacement in nature. In seeking some clue to the phenomena we should note that acetylglucosamine is the basic monomer of chitin, while acetylgalactosamine is a prominent constituent of chondroitin sulphuric acid which is associated with collagen. It is just possible that some change in the nature of the sugars produced determines whether the chitin system or the collagen system will be developed as the basic material of cuticle or tendon.

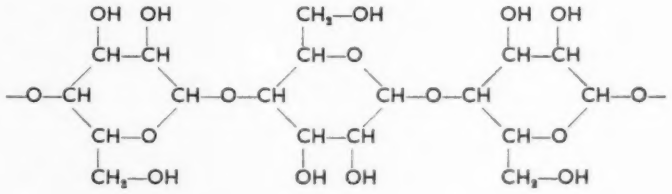
The main types of extended polypeptide chains which have been recognized, and the principal features of intramolecular bonding have now been described. These things are fundamental to the harder problems concerning the various ways in which polypeptide chains can be folded upon themselves. We can at least foresee some of the regular types of folding and these simpler cases have been indicated.

Other natural polymers besides proteins add to our list of fundamental structures. There is special interest in the polysaccharides because of the complexes they form with proteins.

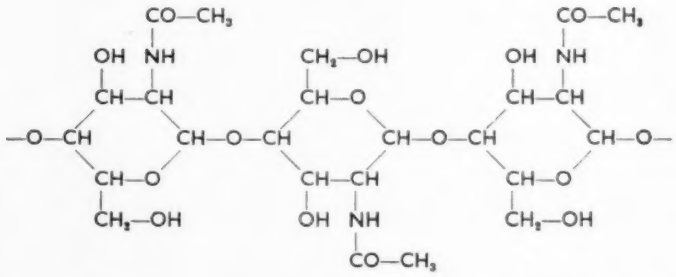
POLYSACCHARIDES

We shall consider here only a few principal points concerning straight chain polysaccharides. The simpler straight chain types *e.g.* cellulose, chitin, alginic acid and pectin are important since structurally they are fairly completely analyzed and they show very clearly certain principles of fibre structure that have undoubtedly helped to give confidence in interpreting more complex natural polymers such as the fibrous proteins.

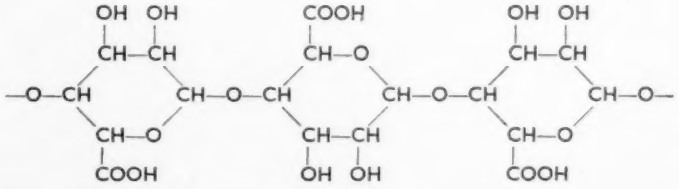
Cellulose



Chitin



Alginic acid and Pectin



In each of the above polysaccharides only one monomer is present, a fact contrasting with the diversity of the units making up a natural polypeptide as in *I*. The polypeptide analogue is to be found in synthetic products like polyglycine, polyalanine *etc* but many natural polysaccharides must exist in which the repeating sugar units are of many different kinds, and thus we can see a basis for their specificity.

Cellulose is the most abundant polysaccharide in nature and virtually

makes possible plant life as we know it. It has been studied most extensively because of its great commercial importance. Chitin is structurally very closely related but is much more abundant in animal structures. It has such an important place in the structure of Arthropods that in a sense it makes this group of creatures possible. The obvious function of cellulose and chitin is as a mechanical support. We may suspect other functions but do not know of them definitely.

Two major interests in the intensive study of these polysaccharides centre around the two questions :

a What reactions do they have with proteins or other biomolecular substances ?

b What is the mechanism of their synthesis and of their secondary fabrication ?

The approach to these questions has been through the study of the purified polymers. For cellulose, *Figure 19*, alginic acid, *Figure 20*, and chitin, *Figure 24*, the crystallography requires an equal number of chains running in opposite directions as in *Figure 21*, which are views of crystal cells down the length of the molecular chains.

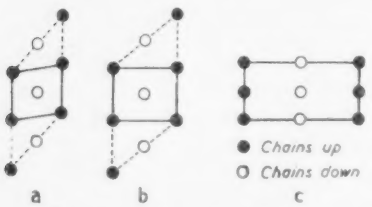


Figure 21. Unit cells of a *cellulose,* b *alginic acid, and* c *chitin*

This sets some limit to what must happen during the deposition of the chains. The simplest view is that when synthesized the chains are comparatively short and capable of free rotation. They could aggregate as short-length crystallites and then be polymerized secondarily into the relatively huge fibrils we see in the electron microscope or optically. On the other hand, if cellulose and chitin are produced as oriented growths from the beginning with great fibre length, then it is likely that an equal number of chains are produced of opposite direction. If intimate association with a protein surface is necessary during the production of these polysaccharides, opposite direction of chain deposition could be readily explained by the opposite directions of polypeptide chains. In a protein monolayer of the β-type it is relatively certain that neighbouring chains run alternatively up and down.

The structure of the fibrils of cellulose and chitin is comparatively well known, but the intramolecular processes occurring between synthesis and fibril formation are much more obscure. Recognition of some of the intermediate steps should take us nearer an understanding of the growth of these fibrils and the primary conditions obtaining during the synthesis of the constituent molecular chains.

The chitin system is of particular interest as it shows well defined

modification of structure in its natural association with protein[10]. This has a bearing on problems of cellulose production on the one hand, and on the protein/polysaccharide complexes of mucopolysaccharides and mucoproteins on the other. Chitin is a condensation product of N-acetylglucosamine which is an important constituent of mucins and occurs as a prosthetic group in proteins. There are some curious things about the internal bonding in chitin. The main symmetry elements in the crystal have been deduced from x-ray fibre photographs of lobster tendon by MEYER and PANKOW and they favour the space group P 2, 2 2, and give cell dimension of $a = 9{\cdot}4$, $b = 10{\cdot}46$, and $c = 19{\cdot}25$ Å.

The cell of *Figure 22* pictures the chains as separated by 9·6 Å in the direction of the plane of the sugar rings, and by *ca* 4·65 Å in a direction perpendicular to this. Chitin is a structure of great chemical resistance and tensile strength and it is only natural to ask questions about the internal molecular linkages. Hydrogen bonds, such as are present in cellulose, are ruled out along the c axis of chitin because of the large chain separation.

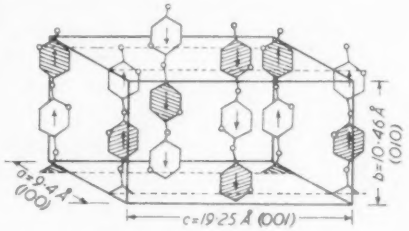


Figure 22. Unit cell of chitin

Because of the aminoacetyl side chains there is the possibility of C=O · · · N—H bonding of the type common in proteins, and this would give a ready explanation of the 4·65 Å separation found both in proteins and in chitin. Chitin has a well defined orthorhombic lattice as is also the case in typical β-proteins.

If we attempt to make C=O · · · N—H bonds between chitin chains according to the lattice of *Figure 22*, we find these are of two different

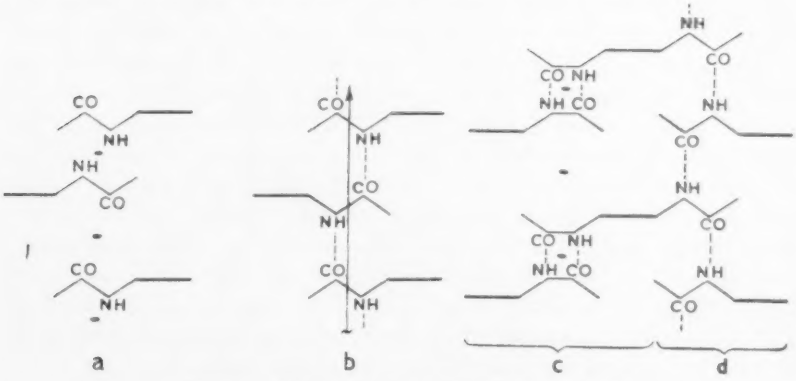


Figure 23. Bonding by C=O···N—H *groups in chitin*

kinds : the cell can be divided into two different piles of chains *viz* those related by a twofold rotation axis as in **a** and those related by a screw axis as in **b**.

Within the pile of chains of the type **b** the ordinary *trans* bonding of CO and NH groups is possible, while in the arrangement **a** there is no C=O · · · N—H bonding unless the side groups rotate and give *cis*-type bonding as in **c**.

In an infra-red study of purified insect cuticle chitin DARMON has found strong absorption corresponding to two types of NH-stretching frequencies which from work by DARMON and SUTHERLAND[35] are believed to correspond to the *trans* and *cis* types of C=O · · · N—H bonding. These appear to be the principal intramolecular bonds in chitin, for there is no evidence of frequent bonding between hydroxyl groups, but the latter type of bonding, which is characteristic of cellulose, does become prominent when chitin is changed to chitosan where half the acetyl and amino groups are removed. The two different piles of chains, as depicted in *Figure 23*, give a possible explanation as to why half the acetyl and amino groups are lost during the formation of chitosan. At present we cannot say whether this loss takes place in the **a** or in the **b** type of pile.

The above observations support the type of lattice which has been derived from x-ray studies. We also learn that the principal interchain bond in chitin is of the same type as the characteristic backbone linkage in proteins. As regards sets of chains this linkage in chitin is probably directed as in *Figure 3***c**.

The structure of purified chitin which we have been considering is present only rarely in natural structures. Where the condition of the natural membrane is fresh and nearest to that in which the substance was synthesized there are marked disturbances in the lattice in planes at right angles to the molecular chain axis. In addition, as in *Figure 25*, it is very striking how the 010 reflection is strongly apparent. Intermediate conditions between this and the effective absence of 010 as in pure chitin, *Figure 24*, are shown by *Phycomyces* chitin and the chitin of the ovipositors of *Sirex gigas*, *Figure 26*. Here there is spreading of 011 along the layer lines to make a streak comprising 011 and 010. In this series, in all probability, we have polymorphic transition between two forms, the one being characteristic of pure chitin and the other characteristic of natural blowfly larval cuticle and a wide variety of other soft cuticles in Arthropods. Typical of the latter is the enlarged spacing along the c axis, *Figure 27*, and this is shown also by the natural *Phycomyces* structure which is compared with purified insect chitin in *Figure 28*. The pure chitin lattice is very stable, but the modified chitin lattice of soft insect cuticles is highly unstable. It changes readily to

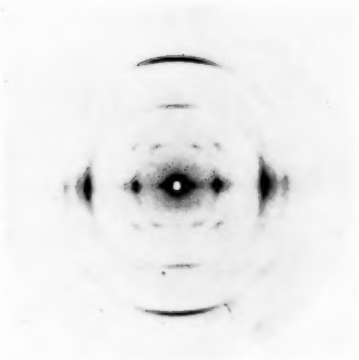


Figure 24. Purified chitin from blowfly larval cuticle

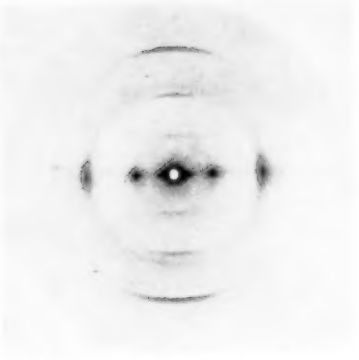


Figure 25. Blowfly larval cuticle before removal of the protein fraction

Figure 26. The ovipositor of Sirex gigas *showing the high orientation of molecular chains parallel to the long axis of the ovipositor*

Figure 27. Line spectra. a *Natural chitin protein complex of blowfly larval cuticle.* b *Purified chitin of blowfly larval cuticle*

Figure 28. Line spectra. a *Natural chitin complex in* Phycomyces *sporangiophores.* b *Purified chitin from insect cuticle*

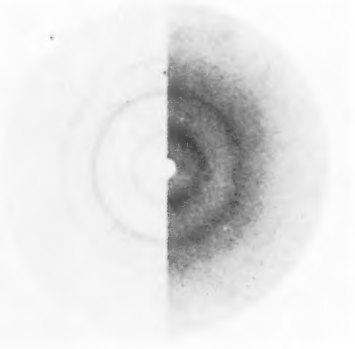


Figure 29. Comparison of purified insect chitin, A, *and insect cuticle protein (arthropodin),* B

the pure chitin lattice when the protein content is altered by heat or by dilute acids and alkalis. By contrast the modified lattice becomes highly stable after fixation of the protein by tanning. The data thus encourage the view that the pure chitin lattice is an end product of degenerative processes, while in the native state of natural membranes the modified chitin lattice is to be interpreted in terms of interaction between protein and polysaccharide. The type of protein occurring in association with chitin is highly water soluble and exists naturally in the β-form. The main dimensions of chitin and this protein are essentially identical, as is evident from the comparison photographs of *Figure 29*. They could form comparatively stable interpenetrating lattices, and one model based on these data has already been considered. But a number of other models are conceivable especially those linking chitin and protein via the C=O · · · N—H type of linkage, or, considered rather from the β-polypeptide grid aspect, main chains could be weakly cross linked by a chitin chain as in *Figure 30*.

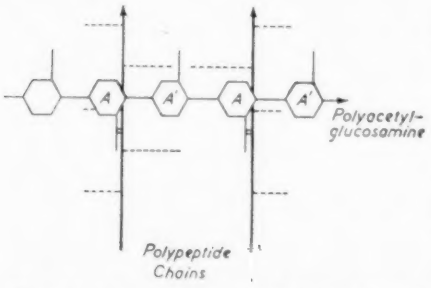


Figure 30. Possible interaction between proteins and polysaccharide chains containing acetylglucosamine residues

This seems a reasonable suggestion to account for the stability of insect cuticle protein when within the cuticle, and its high water solubility and low viscosity when extracted. It offers also a possible model to account for the rather similar properties of hyaluronic acid which, as extracted, is less viscous than when inact with tissue proteins, where it forms a highly viscous system. Hyaluronic acid contains equimolecular proportions of the residues of acetylglucosamine and glucuronic acid. These may be arranged as A A and A′ A′ respectively in *Figure 30*. Whatever may be the final details of these interactions the guiding principles are bound to lie in the forms of the chain molecules concerned and the potentialities of reactive groups.

A remarkable type of polysaccharide chain has been elucidated by ASTBURY's investigation of alginic acid[29, 30], *Figure 21*. This is a condensation product of β-*d*-mannuronic acid, as shown on p 61, and has much the same association with cellulose in certain algae as pectin (poly-α-galacturonic acid) has more generally in plants. And according to present indications alginic acid and pectin have a closely similar molecular structure. The surprise in the alginic acid structure is its departure from the usual configuration of the 1,4 linkage of sugar residues.

According to the diagrams of *Figure 31*, which view the ring of one sugar residue face on in **a** and edge on in **b**, the residues in the cellulose type are linked through oxygen atoms placed at bond positions O_c, O'_c, while in the alginic acid type the linking oxygens are at O_a, O'_a. These alternatives give identity periods of 10·3 Å and 8·7 Å, respectively. This is only the beginning of the story of straight chain polysaccharides. Increasing complexity can arise where both types of linkage in *Figure 31* are used within the same chain. Further complexities arise from spirally coiled chain types and from the branched chain types.

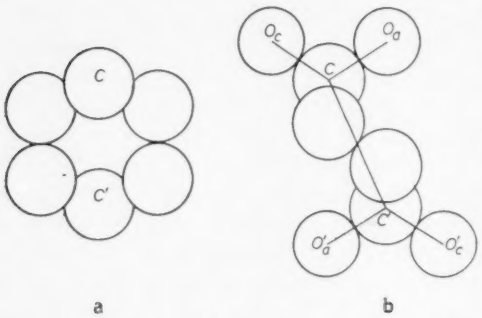


Figure 31. 'Face on' **a** *and 'edge on'* **b** *views of sugar residue from alginic acid or pectin structure*

POLYNUCLEOTIDES

The continuous stream of new information about nucleic acids testifies to the importance of these fundamental structures. Attention is focused on their association with protein to form a self-reproducing system within cells. Genes and viruses appear similar both in principles of structure and in the deeds they perform. Physical studies of nucleic acids establish the basis on which they can be considered as fibrous structures, and suggest the kind of interaction these linear bodies have with fibrous proteins.

The nucleotides which form the unit of structure in nucleic acids are each composed of three kinds of substances, a base (purine or pyrimidine) a pentose sugar (ribose or desoxyribose) and phosphoric acid. These units become polymerized through the phosphoric acid group and give rise to a close-packed pile of nucleotides. In *Figure 32***a** this polynucleotide pile is pictured in name, while in *Figure 32***b** the skeleton of part of the molecule is shown.

This plan of structure is that which the x-ray analysis supports, but it is only with sodium thymonucleate that satisfactory diffraction patterns have been obtained. Sodium thymonucleate films can be drawn into fibres in which the chain molecules are highly oriented parallel to the direction of stretching ; the axial repeat of pattern gives an intense reflection of spacing 3·3 to 3·4 Å, *Figure 33*. ASTBURY's interpretation is that this represents the distance between successive flat nucleotides closely packed like a pile of pennies[31]. A vast amount

of additional information on the diffraction patterns has already been obtained. Whereas much of this remains incompletely understood, it is already clear that along the fibre axis of *Figure 33* there are much larger periodicities which doubtless have their origin in unique arrangements of nucleotides, thus giving a basis for the specificity of

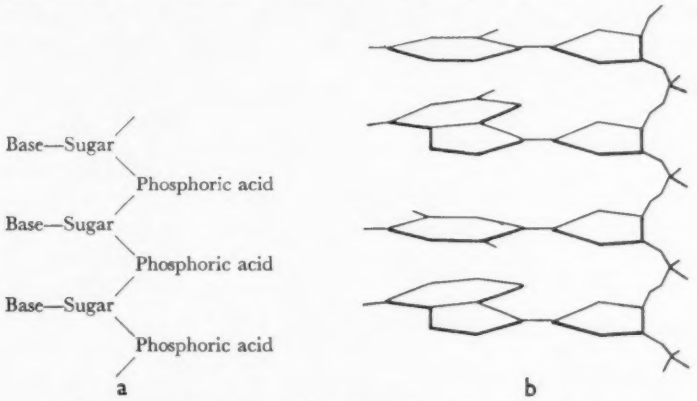


Figure 32. **a** *Composition of a polynucleotide and* **b** *molecular skeleton*

nucleic acids. Sodium thymonucleate has a well defined micellar structure due to the regular side by side aggregation of the columns. The electron microscope has revealed long thin fibrils deposited from dilute solutions of viscous thymonucleic acid, and these are almost certainly equivalent to the micelles recognized by x-rays.

In one sense the above data provide only further quantitative information about a polymer, but this becomes particularly vital when considering nucleic acids in association with proteins. The basic idea about the physical details of this association is that nucleotides and amino acids are both distributed at intervals of *ca* 3·4 Å along the chains of polynucleotides and β-proteins respectively. Thus where nucleic acid and protein chains are brought together groups could interact without distorting either kind of chain. The nucleic acid column is saturated with reactive groups, for example, acidic phosphoric groups on the axis of the column and basic $—NH_2$ groups of the superposed nucleotides. The salmon sperm protein, clupein, is particularly rich in basic groups due to its high arginine content, so that interaction between nucleic acid and clupein should produce a particularly high concentration of salt-like cross linkages. Tough membranes are produced by the above interaction, and an x-ray fibre photograph of this nucleoprotein is illustrated in *Figure 34*. The basic structure of the nucleic acid column is still clear but there are obvious alterations in the side to side arrangement of chains, which can be

appreciated when this diagram is compared with that of the sodium thymonucleate, *Figure 33*. As the pattern of *Figure 34* has not been extensively analyzed we cannot say how neatly the nucleic acid and protein chains fit on one another, but we do know that the columnar nature of the nucleotides is not disturbed.

The simplest association of nucleotide column and β-protein chain would be in a 1 : 1 ratio between nucleotide and amino acid residues. Chemical analysis of the artificial clupein thymonucleate gives a ratio of 2 : 3, a figure which corresponds closely to an equivalence of nucleic acid acidic groups and clupein basic groups[32]. To account for this, various distortions of the clupein chains have been suggested taking us away from the concept that the similar spacial distribution of nucleotides and β-chain amino acids is of special significance. Further, reports that the fibrous product of combination between nucleic acid and edestin, which contains many acidic groups in addition to basic groups, gives a diffraction pattern in which neither nucleic acid structure nor protein structure can be recognized, serve to emphasize how both polymers may be greatly distorted from the straight chain form as a result of interaction.

We can proceed a little further with the structure of thymonucleic acid and nucleoprotein. x-ray diffraction patterns of whole thymus gland of the mouse reveal unmistakably the strong repeat of pattern of 3·3 to 3·4 Å, *Figure 35*. This pattern is of dried, lightly flattened gland. Where attempts were made to produce higher orientation, by double compression in two directions at right angles, certain changes of structure occurred. In *Figure 36* the nucleic acid period 3·3 to 3·4 Å has increased in intensity and improved in orientation, but a distinct pattern of the β-protein type has also appeared. This latter may result from a transformation of folded polypeptide chains to the extended β-form, or it may result from the separation of protein which was in combination with nucleic acid. Similar compression of liver tissue did not produce this distinct β-type protein, which appears to be a special manifestation of conditions in the thymus gland. It is, however, satisfying to know that the typical nucleic acid columns exist within the intact gland as a dominating structure and that there is a considerable protein fraction capable of existing in the fully extended β-form.

No simple explanation has been found of the self-reproducing properties of nucleoproteins nor of the much more complex affair of how they determine all kinds of changes in the structure and function of the host ; but some understanding of the structure of nucleic acids and of protein chains has been achieved and we can see how these may become associated in the performance of their duties.

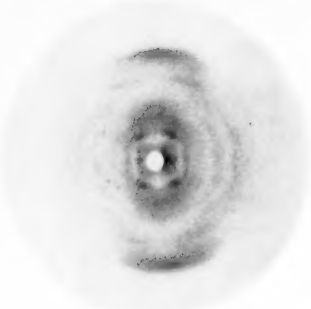


Figure 33. Fibres of sodium thymonucleate

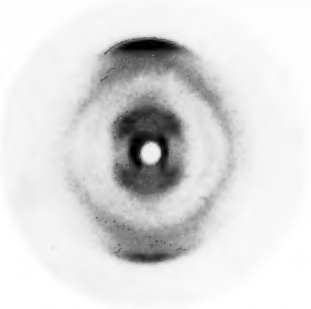


Figure 34. Partly oriented film of clupein thymonucleate

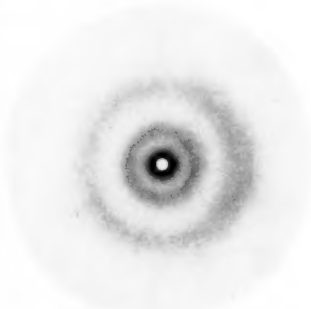


Figure 35. Whole thymus gland of the mouse with the X-ray beam parallel to the flattened tissue

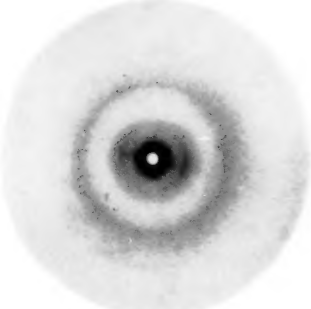


Figure 36. The same as in Figure 35 but more completely oriented by compression. Shows the development of the typical β-protein pattern superposed on that of the nucleic acid structure

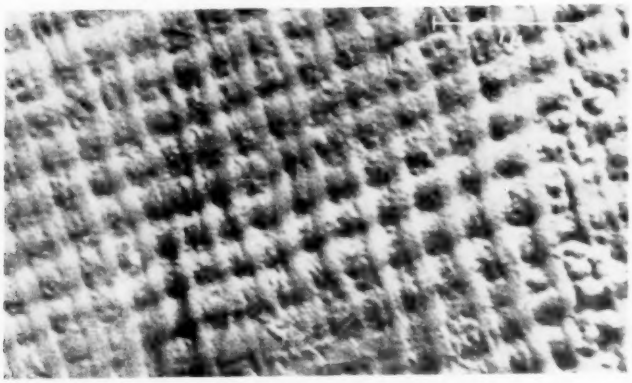
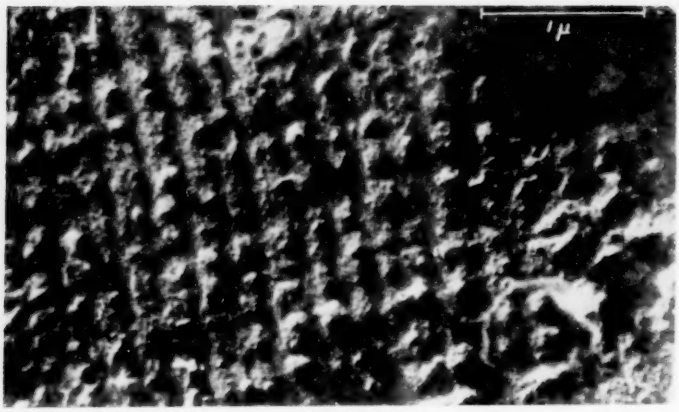
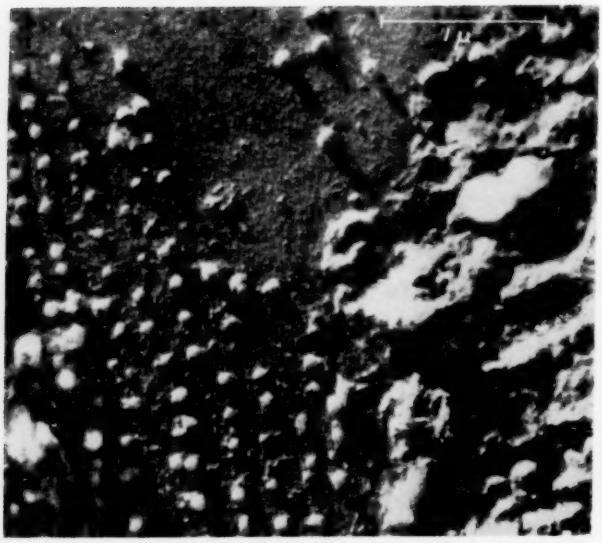


Figure 37. The crossed fibrils of the two lowest fibre layers of the earthworm cuticle; electron micrograph, shadow cast with chromium





Figures 38 and 39. Transition layers between the epithelial cells and the formed fibril layers of the earthworm cuticle

Many studies have been devoted to structures with well defined molecular orientation because these conditions seemed likely to reveal important things about the way in which biomolecular substances are organized during growth. Regular orientations are frequently shown by natural fibrous substances, both polypeptide and polysaccharide in nature. The principal kinds of orientation are :

1 the alinement of molecular chains parallel to a given axis

2 orientation of a crystal axis perpendicular to a natural surface

3 the regular crossed fibre directions in natural membranes, commonly spoken of as crossed lamella systems

Mechanical forces— In certain cases the first two types are satisfactorily explained in terms of simple mechanical forces. For example, tension in a given direction produces parallel orientation of molecular chains and this is the explanation of the parallel orientation in natural silk fibres or the transverse orientation of chitin chains of the blowfly larval cuticle during pupal contraction[33]. Here mechanical forces are wholly responsible for the orientation of the chains.

The second type of orientation can be produced in experimental material in several ways. The simplest is to allow a wet membrane to dry and then the preferred orientation occurs spontaneously in certain materials, such as bacterial cellulose and soft insect cuticle systems. In a wider variety of materials a similar effect is achieved by pressing a membrane between glass plates. Another somewhat different procedure is to apply lateral tension to a membrane, and then the longer of the two smaller axes may become parallel to the plane of the surface. Thus where there is stretching of a membrane, or internal pressure due to turgor, preferred orientation of crystallites with reference to the plane of the surface is likely to occur.

Intermolecular forces— The most interesting cases are those where no simple mechanical explanation is available and indeed the explanation may be very complicated. The phenomena may be broken down into the properties of particles or single molecules. Even if these are nearly globular they will often be polarized in structure and reactivity. Interaction between such particles will lead to polymerization in only one direction, or mainly in one direction with side to side aggregation as well. The classic example is fibrin, which produces needle-like growths visible in dark field or ultramicroscope ; an older example is the reprecipitation of collagen, while the electron microscope has revealed very numerous cases of a similar kind. Intermolecular forces are responsible for the orientation in these cases, sometimes involving enzymes as in the fibrinogen-fibrin change, and sometimes weaker

cohesive forces as in the production of fibrous actin and fibrous tropomyosin. There is no difficulty then in appreciating the general nature of the production of an oriented fibril in isolation, but to explain, for example, why a whole muscle fibre has a uniform parallel orientation of molecular chains is much more difficult. Some factor is involved which controls the orientation of all the particles over a large area. The long range molecular forces, such as hold a dilute solution of tobacco mosaic virus in a high degree of orientation, are a possibility[32]. What determines the axis along which these linear fibrils extend is the basic problem in muscle fibril orientation and in numerous other instances as well.

The problem of the orientation of fibrils in crossed lamella systems is even more complicated, but some progress has been made here concerning the factors involved. The development of the extracellular cuticle in annelids has been studied extensively with the electron microscope[34]. The basic structure here is of the collagen type. In the earthworm the completed cuticle consists of a number of layers of parallel fibrils, the direction of the fibrils lying at right angles in alternate layers. The production of these layers at the surface of the epithelial cells may be studied and information obtained on two principal points *viz* the processes by which fibrils are formed and the manner in which the pattern is determined. More general aspects have been described elsewhere[34], but the three electronmicrographs, *Figures 37, 38* and *39*, show main features of the development of the cuticle pattern. *Figure 37* shows the two lowermost layers of the cuticle proper *i.e.* where definite fibrils are already formed, while *Figures 38* and *39* show the predominantly granular layers between the cuticle proper and the epithelial cells. The conspicuous granules of *Figure 39* indicate that the crossed pattern is determined at the cell surface by the production or adsorption of these particles at regularly arranged sites. Often the obvious granules, which are about 0·1 μ in diameter, are clearly subdivided into a large number of much smaller corpuscles. *Figure 38* represents an intermediate stage between those of *Figures 37* and *39*, and shows the development of ridges in a preferred direction corresponding to one of the fibre layers.

The figures suggest many complex phenomena. Primarily they indicate the presence of minute corpuscles as the fundamental unit of the fibrils. The actual width of the fibrils seems to be determined by the aggregation of the smaller corpuscles into larger discrete granules. These larger bodies are arranged in a regular array which gives the basis for the crossed lamellar system. In one layer they become alined along a specific direction, and according to *Figure 38* this seems to occur by the disaggregation of the component corpuscles of the large

granules. The determination of this preferred direction may lie in the particular polarization of that layer of granules, and that of the next successive layer by a polarization in a direction at right angles. These studies bring out some of the complex phenomena involved in orientations that occur over large areas. The earthworm cuticle may be described as an open network of comparatively wide fibrils. A limiting structure can be envisaged where the smallest corpuscles are in close packed array of the hexagonal type or in square net array. Polymerization along rows of corpuscles in these arrays would lead to parallel orientation of narrow fibrils closely packed side by side.

Molecular chains and biological problems— Nearly the whole of living matter is included in the range ' chain molecules to visible fibres and membranes'. We aim to discover how the chain molecules are produced, often with their diverse components accurately arranged according to countless different plans. How do these first fibres become arranged neatly within a protein corpuscle for example, or within a protein fibril ? Quest for knowledge of these most magnificent of all molecules is bound up with details of the structure and form of polypeptide chains. Interactions of polypeptides, polynucleotides, polysaccharides and lipoids, as well as all the secondary fabrications of chain molecules, are other things we have to learn about in order to understand living matter. These structural aspects carry the study of anatomy to its foundations. Those who work from the field of optical microscopy through the new knowledge that the electron microscope provides come into ever closer contact with the fundamental structures we have considered. Real knowledge of these things will illuminate the whole path that has been covered.

From the functional aspects we are interested in how things go wrong within the realm of our fundamental structures, or how they are modified in the course of evolution. Many a biological phenomenon is an event of chains. Well emphasized among these is the concept of the polypeptide chain as a dynamic system capable of a series of transformations in the process of which it does work. Here we are closing in on the long mystery of muscular movement ; and we may also throw light on many other types of motion in biological systems.

REFERENCES

[1] Astbury, W. T. and Street, A. *Phil. Trans roy Soc.A* 230 (1931) 75
[2] — and Woods, H. J. *ibid A* 232 (1933) 333
[3] — *Trans. Faraday Soc.* 29 (1933) 193
[4] — *ibid* 36 (1940) 871
[5] — *Nature, Lond.* 163 (1949) 722
[6] Cohn, E. J. and Edsall, J. T. *Proteins, Amino Acids and Peptides* New York, 1943
[7] Astbury, W. T., Dalgliesh, E., Darmon, S. E. and Sutherland, G. B. B. M. *Nature, Lond.* 162 (1948) 596

[8] AMBROSE, E. J., ELLIOT, A. and TEMPLE, R. B. *Nature, Lond.* 163 (1949) 859
[9] ASTBURY, W. T. *Chem. and Industr.* 60 (1941) 491
— *J. chem. Soc.* (1942) 337
[10] FRAENKEL, G. and RUDALL, K. M. *Proc. roy Soc.B* 134 (1947) 111
[11] HUGGINS, M. L. *Chem. Rev.* 32 (1943) 195
[12] MARTIN, A. J. P. *Symposium on Fibrous Proteins*, Soc. Dye. Col. (1946) 1.
[13] ASTBURY, W. T. and DICKINSON, S. *Proc. roy Soc.B* 129 (1940) 307
[14] — *ibid B* 134 (1947) 303
[15] WHEWELL, C. S. and WOODS, H. J. *Symposium on Fibrous Proteins*, Soc. Dye. Col. (1946) 50
[16] RUDALL, K. M. *ibid* (1946) 15
[17] ASTBURY, W. T., DICKINSON, S. and BAILEY, K. *Biochem. J.* 29 (1935) 2351
[18] — *Nature, Lond.* 137 (1936) 803
[19] — and BELL, F. O. *Cold Spring Harbor Symp. Quant. Biol.* 6 (1938) 109
[20] JAKUS, M. A. and HALL, C. E. *J. biol Chem.* 167 (1947) 705
[21] BAILEY, K. *et alii Biochem. J.* 43 (1948) 271, 279, 282
[22] PERUTZ, M. F. *Proc. roy. Soc.A* 195 (1949) 474
[23] ASTBURY, W. T. and LOMAX, R. *Nature, Lond.* 133 (1934) 794
[24] — and MARWICK, T. C. *ibid* 130 (1932) 309
[25] RUDALL, K. M. *Biochim. Biophys. Acta* 1 (1947) 549
[26] ASTBURY, W. T., PERRY, S. V., REED, R. and SPARK, L. C. *ibid* 1 (1947) 379
[27] — *Nature, Lond.* 160 (1947) 388
[28] — *J. int. Soc. leath. Chem.* 24 (1940) 69
[29] — *Essays on Growth and Form* Oxford, 1945
[30] — *Nature, Lond.* 155 (1945) 667
[31] — *Symp. Soc. Exp. Biol.* 1 (1947) 66
[32] BERNAL, J. D. and FANKUCHEN, I. *J. gen. Physiol.* 25 (1941) 111
— — *Trans. Faraday Soc.* 42B (1946)
[33] FRAENKEL, G. and RUDALL, K. M. *Proc. roy Soc.B* 129 (1940) 1
[34] REED, R. and RUDALL, K. M. *Biochim. Biophys. Acta* 2 (1948) 7
[35] DARMON, S. E. and SUTHERLAND, G. B. B. M. *Nature, Lond.* 164 (1949) 440
[36] MIZUSHIMA, S. *et alii ibid* 164 (1949) 918

3

SCATTERING OF VISIBLE LIGHT AND X-RAYS BY SOLUTIONS OF PROTEINS

G. Oster

THE size, shape, and relative positions of protein molecules may be determined by visible light scattering and by x-ray scattering. These parameters may be determined when the protein is subjected to any physical or chemical condition and may conveniently be studied as a function of time. From the variation in relative positions of the particles with protein concentration, the equation of state of the solution (the osmotic pressure as a function of concentration) and, consequently, the nature of the interparticle forces may be deduced. Besides being of interest in their own right, protein solutions may serve as possible models of biological structures ; for solutions of certain proteins such as those of haemoglobin and of tobacco mosaic virus, they appear to be similar to the actual state of these proteins as they exist in living cells. It should be possible to determine the size, shape and relative positions of macromolecules in living cells and tissues by means of scattering but many technical difficulties in such measurements must first be overcome. In any case, the cytologist must be aware of such scattering effects when critically evaluating optical observations made on biological structures. In this chapter a summary of the results of the theory of scattering will be given and will be illustrated by examples with protein solutions.

The scattering by matter of visible light and of x-rays arises from the same origin. The electric field associated with the incident radiation induces periodic oscillations in the material. The material then serves as a source of secondary radiation and scatters radiation with a wave length equal to that of the incident radiation. The relatively small amount of radiation which is emitted with altered wave length (*e.g.* Raman effect, Compton effect, and fluorescence) will be neglected. The dependence of the total intensity of scattering on wave length for visible light is different from that for x-rays but for protein solutions, at least, the angular distribution of scattered radiation will be identical, for a given particle size to wave length ratio for the two cases. This arises since with all substances the index of refraction for x-rays is very close to unity and for visible light the index of refraction of proteins relative to water is nearly unity (about 1·15), from electromagnetic theory it follows that the angular distribution of scattering

will be determined by the geometrical arrangement of the scattering elements, treated as dipole scatterers, making up the protein particle. If the system consists of many particles not randomly arranged with respect to one another, the interferences between the scattered wavelets from each of the particles must be considered.

The formal identity of the theory of visible light and of x-rays is convenient since it allows a direct comparison of the results obtained from experiments using radiations which differ in wave length by the order of one thousand. The problem of scattering naturally divides itself into four categories depending on whether the particles are small or large compared with the wave length and whether or not they occupy independent positions in space relative to one another.

SMALL INDEPENDENT PARTICLES

By small particles we mean here particles whose largest linear dimension is about one tenth the wave length of the radiation. Since we are dealing with proteins, this condition is satisfied by using visible light as the radiation and the particles are smaller than 50 mμ, corresponding for spherical proteins to a molecular weight of about 50 million. The condition of independence of the particles is satisfied if the solution is thermodynamically ideal.

For small optically isotropic scatterers the intensity of scattering is proportional to $1 + \cos^2\theta$ where θ is the angle between the observer and the proceeding incident beam so the scattering is symmetrical about 90° to the incident beam. When a beam of light traverses a light scattering system its intensity is decreased by virtue of the energy withdrawn from the beam in the form of scattered radiation. For colourless particles the logarithm of the fractional decrease in the transmitted intensity I is given by the turbidity τ where $I = I_0 \exp(-\tau l)$, I_0 being the initial intensity and l the path length of the scattering system. Application of RAYLEIGH's law[1] for independent point scatterers in solution gives[2]

$$\tau = HcM$$

$$H = \frac{32\pi^3 n_0^2}{3N\lambda^4}\left(\frac{n - n_0}{c}\right)^2 \qquad \ldots\ldots(1)$$

where c is the weight concentration in gm protein per ml, M is the molecular weight, λ is the wave length of the incident light, N is Avogadro's number, n_0 is the index of refraction of the medium, and n is the index of refraction of the solution. For protein solutions the index of refraction increment $n - n_0/c$ is independent of the concentration and is equal to about 0·17.

There are a wide variety of experimental arrangements for making

light scattering measurements on systems of low turbidity. In general one needs a strong monochromatic source of light *e.g.* a high pressure mercury lamp with suitable coloured filters, and a very sensitive device at right angles to the beam for determining the scattered intensity such as a photomultiplier tube or, if a matching is made of the scattered intensity with the intensity of a portion of the incident beam, the dark-adapted eye is suitable. The incident beam may be convergent with its focus at the centre of the scattering sample or may be rendered parallel with suitable lenses. The system must be rendered free from stray radiation by coating the surfaces of the scattering cell and the lenses to reduce reflections from the incident beam and the scattering cell and detector must be protected from outside light. The absolute turbidity of the sample may be determined by comparing the intensity of scattered light with that for a liquid of known turbidity such as carbon disulphide.

If the solution is sufficiently turbid the absolute turbidity may be measured directly in a conventional colorimeter or spectrophotometer.

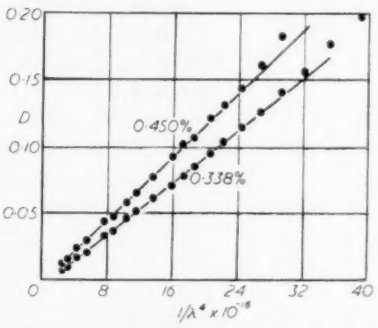


Figure 1. Optical density as a function of inverse fourth power of the wave length for bushy stunt virus in water

In *Figure 1* are shown the results obtained with a spectrophotometer for dilute solutions of tomato bushy stunt virus[3]. The turbidity, given in terms of the optical density D (for a one centimetre path length $\tau = 2{\cdot}303D$) obeys Rayleigh's inverse fourth power of the wave-length law over the visible range, but at shorter wave lengths there is real light absorption due to the presence of chemical constituents absorbing in the ultra violet. The molecular weight of the virus particles (spherical particles about 30 $m\mu$ in diameter) calculated from the slopes of the curves is in agreement with that determined by other methods. The turbidity of the pure solvent is automatically taken into account when the turbidity of the solvent is taken as the zero reading on the spectrophotometer. The light scattering method gives the molecular weights of the dry protein, since the water of hydration of the protein has nearly the same refractive index as has the solvent water.

PUTZEYS and BROSTEAUX[4] were the first to use light scattering to determine the size of proteins. They found that the intensities of scattering from solutions of ovalbumin, amandin, excelsin, and haemocyanin were directly proportional to the molecular weights of these

proteins determined by other physico-chemical methods. In recent years the absolute molecular weights of some of the smaller proteins (ovalbumin[5], rabbit antiserum[6] and serum albumin[7]) have been determined by light scattering and the molecular weights found are within 10 per cent agreement with those found by other methods. These protein solutions exhibit no depolarization *i.e.* the light scattered at 90° is perfectly plane polarized indicating that the particles are optically isotropic.

Equation 1 is the scattering for a monodispersed system. For a polydispersed system of ν_j (per unit volume) small isotropic particles of volume v_j, equation 1 becomes

$$\tau = A\sum_j \nu_j v_j^2 \qquad \ldots.(2)$$

where A is a function of the wave length and of the index of refraction increment (note that $\nu = cM/\mathcal{N}$ and $M \sim v$). Thus, for a mixture of different size particles the turbidity is weighted in the direction of the larger size particles. It is necessary, therefore, if the scattering from the smaller particles is to be studied, to clarify the sample, preferably by centrifuging, to remove the larger particles. Equation 2 has been used to obtain expressions for the turbidity as a function of time for kinetic processes[8]. The variation of the turbidity with time can give an indication of the type of polymerization or degradation process occurring in the system and the rate constants of the elementary steps in the process. Examples of such processes studied by time rate of change of the turbidity are given by observations on the rate of salting out of bushy stunt virus[8], the precipitation of ovalbumin and its antiserum[8], and the slow aggregation of ovalbumin[9].

SMALL NON-INDEPENDENT PARTICLES

For non-ideal solutions the thermal motions of the particles are not independent and, therefore the total intensity of scattered light will not be the mere summation of the intensities from the individual particles. Some destructive interference will occur and result in a decrease in intensity of scattered light from that expected for the same concentration of independent particles. The phase relations can be computed in terms of the radial distribution (the probability of finding a particle at any distance from a reference particle) and this will be done in the last section of this chapter. However, the scattering for a non-ideal solution of small particles may, by the application of the theory of fluctuations, be directly related to the osmotic pressure of the solution. According to this theory the particles in solution are not uniformly distributed in space, but rather, in any small volume of the solution the number of particles varies from instant to instant as a

consequence of their thermal motions. The local fluctuations in density of particles will give local inhomogeneities in refractive index and will scatter light according to the SMOLUCHOWSKI-EINSTEIN equation [10, 17] (for simple derivation see OSTER[2])

$$\tau = H\frac{RTc}{\partial P/\partial c} \qquad \ldots\ldots(3)$$

where R is the gas constant, T the absolute temperature, and P the osmotic pressure of the solution. The osmotic pressure of a solution may be written as a power series in c which for dilute solutions becomes

$$P = \frac{cRT}{M} + Bc^2 \qquad \ldots\ldots(4)$$

where B is a measure of the deviation from ideality. Substituting equation 4 into equation 3 we obtain

$$H(c/\tau) = \frac{1}{M} + \frac{2Bc}{RT} \qquad \ldots\ldots(5)$$

Thus, at low concentrations a plot of $H(c/\tau)$ against c gives a straight line whose intercept is the reciprocal of the molecular weight and whose slope is determined by B. At infinite dilution ($c \to 0$) equation 5 reduces to equation 1, the case for independent particles. At high concentrations the higher terms in the power series for P become important and the $H(c/\tau)$ against c curves upwards.

For protein solutions B will be determined only by the shape and the interactions of the particles ; the more elongated the particles and the greater their mutual repulsions the greater the value of B. Formally B also contained a term expressing the interaction of the particles with the solvent. For proteins, however, the heat of interaction (heat of dilution) is not important except at extremely high protein concentrations. That is, when a small amount of water is added to a dry protein there is vigorous interaction of the charged groups of the protein with the water (water of hydration) with a consequent evolution of heat. On further dilution, however, there is no appreciable heat evolved. For elongated particles such as tobacco mosaic virus solutions, for example, light scattering studies[11] show that B is greater the lower the electrolyte concentration *i.e.* the greater the thickness of the repelling ionic double layer around the particles. When sufficient electrolyte is added to reduce the thickness of the layer, the value of B observed is consistent with that calculated for a rod 280 mμ in length and 15 mμ as is observed by other methods. For proteins B should be a minimum at the iso-ionic point where the electrostatic repulsive force between the particles is a minimum. EDSALL, EDELHOCH, LONTI and MORRISON[34] have found this to be the case for

solutions of bovine serum albumin. They further found that the minimum in B shifts to more acid pH values for higher sodium chloride concentrations indicating binding of chloride ions by the protein.

The light scattering method of determining the osmotic pressure is particularly useful in the case of large particles such as viruses whose osmotic pressure is extremely difficult to measure by conventional methods. Even for small protein molecules the light scattering method is useful since one may work with low electrolyte concentrations where Donnan membrane corrections for an osmometer render the latter method useless.

LARGE INDEPENDENT PARTICLES

Particles which have a linear dimension comparable or greater in size than the wave length of the incident radiation will scatter more light in the forward than in the backward directions. The theoretical angular distribution of scattering is given by the solution of the electromagnetic problem of a plane wave falling on a particle with index of refraction close to that of the medium in which it is imbedded. For a sphere of diameter L and of uniform scattering material within it the relative scattering F^2 (the intensity is taken as unity at $\theta = 0$) is given[12] by

$$\left[\frac{3}{z^3}(\sin z - z\cos z)\right]^2 \qquad \ldots.(6)$$

where $z = 2\pi L/\lambda'$. $\sin(\theta/2)$ and λ' is the wave length of the radiation in the medium. For x-rays, λ' is the same as the wave length of the incident radiation but for visible light $\lambda' = \lambda/n_0$, where n_0 is the visible light refractive index of the medium in which the particles are located.

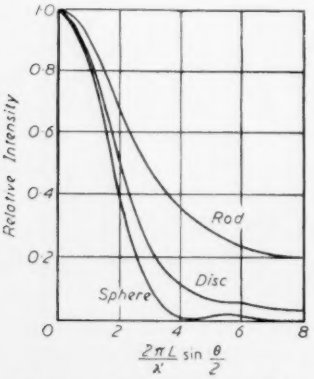


Figure 2. F^2 versus z for rod, disc, and sphere

Equation 6 must also be multiplied by the polarization term symmetrical about 90°, namely, by $\frac{1}{2}(1 + \cos^2\theta)$. That is, superposed on this symmetrical curve is a curve (equation 6) giving greater scattering in the forward directions ($\theta < 90°$). In the limiting case of very small particles ($L/\lambda << 1$) F^2 is nearly unity for all z. In the limiting case of extremely large particles ($L/\lambda >> 1$) the angular distribution of scattering becomes that calculated from geometrical optics[2].

Figure 2 shows curves for F^2 as a function of z for spheres (of diameter

L), thin rigid rods (of length L), and thin circular discs (of diameter L). The relative scattering is reduced to one half the value it has at $\theta = 0°$ at z equal to 1·76, 2·05, and 2·80 for spheres, discs and rods respectively. Using x-rays (say the $K\alpha$ line of copper, $\lambda = \lambda' = 1{\cdot}54$ Å) the intensity will be reduced to one half at $\theta = 0{\cdot}7°$ for a spherical protein of diameter $L = 75$ Å ($M \simeq 175{,}000$). Using visible light (say the green line of mercury, $\lambda = 546$ mμ, $\lambda' = 410$mμ), the intensity will be reduced to one half at $\theta = 45°$ for a spherical protein of diameter $L = 130$ mμ ($M \simeq 600$ million). Thus, for x-ray scattering of proteins, measurements must be made extremely close to the transmitted beam. This difficulty has been overcome in recent years by the use of high-powered x-ray tubes with large specimen to photographic plate distance, bent crystal focusing arrangements, and double crystal spectrometers. Small angle x-ray apparatus is reviewed by RILEY[13]. For visible light scattering the angles of observation are large and the angular distribution may be detected by a photomultiplier tube mounted on an arm capable of rotation about a vertical axis[14, 15].

In *Figure 3* is shown the angular distribution of light scattering for dilute solutions of tobacco mosaic virus[2]. *Figure 3*a shows the relative intensity of scattering of green light as a function of angle for a sample of the virus which had been rendered nearly biologically inactive by having been subjected to strong sonic irradiation. This sample is shown by direct examination in the electron microscope[16] to consist of rod-shaped particles one half the length or shorter than the normally biologically active virus ($L = 280$ mμ). The curve of *Figure 3*a shows only a slight difference in angular scattering from the symmetrical polarization curve $\frac{1}{2}(1 + \cos^2\theta)$; for x-rays the polarization term is nearly unity since θ is close to zero. For the normal virus, however, *Figure 3*b, the intensity is twice as great at 45° as that at 135°. The angular distribution which

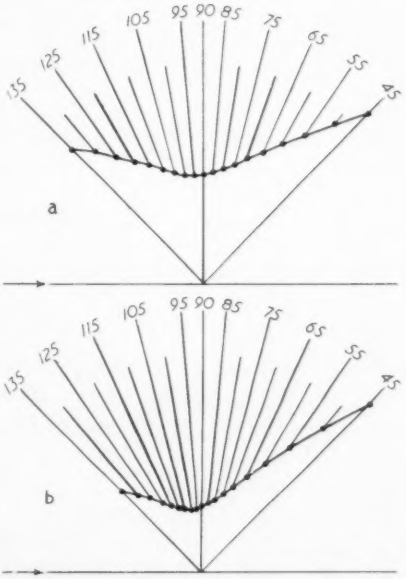


Figure 3. Angular scattering ($\lambda = 546$ mμ) *for solutions* (0·01 *gm*/100 ml *buffered at* p*H* 7·0) *of* a *tobacco mosaic virus which had been subjected to sonic irradiation for* 64 *min, and* b *normal freshly purified tobacco mosaic virus*

is superposed on the polarization curve follows the course of the curve for rods given in *Figure 2*.

DEBYE[17] has pointed out that for particles of low relative refractive index intensity measurements made at two equal angles on either side of 90° allows for a determination of the particle size. Curves of the dissymmetry of scattering *i.e.* the ratio of intensity in a forward direction to that in a backward direction, may be constructed from the curves of *Figure 2* by keeping θ fixed and varying L/λ'. In *Figure 4* are shown the dissymmetry as a function of L/λ' ($\theta = 45°$ and 135°) for spheres, discs, and rods. The dissymmetries at these angles for the sonic irradiated and normal virus particles of *Figure 3* are 1·20 and 2·06 corresponding from *Figure 4* to rods of lengths 103 mμ and 275 mμ respectively. If the density is known the dissymmetry for a sphere gives the molecular weight of the particle since it determines its diameter. For rods or discs, however, the dissymmetry measurements give the length of the largest dimension and the smaller dimension is deduced from the molecular weight determined from turbidity measurements which have been corrected for the dissymmetry *e.g.* measurements on tobacco mosaic virus[18]. The dissymmetry curves approach asymptotes for larger values of L/λ', namely, at $\theta = 45°$ and 135°, 2·44 for rods, 5·98 (small oscillations about this value) for discs. Thus, the dissymmetry characterizes the shape of the particles. The change in shape of the long fibrillar actomyosin molecules on the addition of adenosine triphosphate (ATP) has been studied in this manner[19]. The dissymmetry method may be used to study the kinetics of end to end aggregation of proteins, as has been done for tobacco mosaic virus[11], or the kinetics of breakdown of fibrillar molecules.

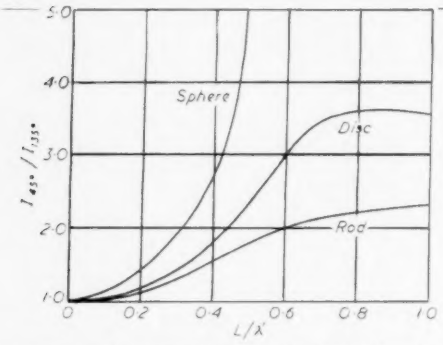


Figure 4. Dissymmetry as a function of L/λ' for rod, disc, and sphere

The size and shape of smaller proteins which show a dissymmetry equal to unity with visible light may be determined by small angle x-ray scattering. Some x-ray scattering studies of a preliminary nature have been made on dilute solutions of ovalbumin[20], edestin[21], insulin[21], tobacco mosaic virus[21], haemocyanin[21], actomyosin[22] and haemoglobin[23]. Of these the experimental data has been analyzed in more detail in the case of haemoglobin[23]. As GUINIER[20] has pointed out, the x-ray scattering immediately around the central beam *i.e.* for

small z, will follow an error curve with a curvature determined by the radius of gyration of the particle. His results for haemoglobin[23] are consistent with the size determined from an analysis of the x-ray diffraction by crystalline haemoglobin[24]. Besides its simplicity, the small angle x-ray scattering method has the obvious advantage that it may be used to determine the size and shape of proteins which are not available in crystalline form.

The wave length dependence of the energy withdrawn from the beam due to scattering gives an indication of the size and shape of the particles. For light scattering from large particles the $1/\lambda^4$ relation (equation 1) will no longer be obeyed except for very large λ as is illustrated in the case of the spherical influenza virus particles ($L \simeq 100$ mμ), *Figure 5*. The total scattered intensity is calculated by integrating the relative intensities (times the polarization factor) over a large spherical region surrounding the particle and multiplying the result by the scattering for a point dipole scatterer. For light scattering the total intensity is proportional to $1/\lambda^2$ and $1/\lambda^3$ for spheres and rods respectively and for x-ray scattering proportional to λ^2 and λ for spheres and rods respectively.

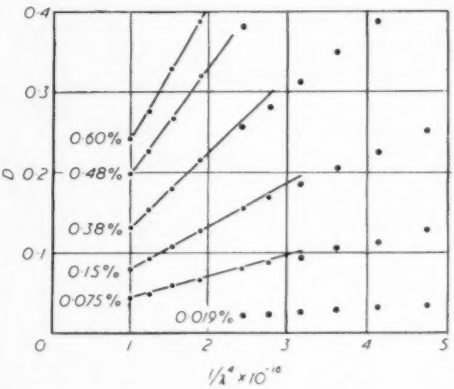


Figure 5. Optical density as a function of inverse fourth power of the wave length for influenza virus in water[3]

NON-INDEPENDENT LARGE PARTICLES

The angular distribution of scattering on a system of large non-independent particles will be determined not only by the finite size of the particles as described above but also by the relative positions of the particles in space. Even for dilute solutions the particles, by virtue of their finite size, cannot occupy all positions in the solution but only those positions not excluded by the particles themselves or more strictly, by a volume given by the sum of the covolume of the particles and their repelling ionic double layers. The system is thus non-ideal and the osmotic pressure will be a power series in the concentration which for very dilute solutions is given by equation 4.

An equivalent approach to the problem is that given by the radial distribution treatment used in the theory of x-ray scattering of non-

ideal gases and liquids. For a very dilute system of spheres of diameter L^* ($L^* > L$ if there is repulsion between the particles) the probability of finding a particle at distance r from any reference particle is nearly unity for $r > L^*$ and zero for $r \leqslant L^*$. This gives a linear decrease in relative intensity with concentration[25] according to the relation

$$F^2(z)\left[1 - \frac{4}{3}\pi\nu L^{*3}F(2z^*)\right] \qquad \ldots\ldots(7)$$

where F is a function given by the square root of equation 6, z^* is the corresponding z for L^*, and ν is the number of particles per unit volume. It may be shown from statistical mechanics that $\frac{4}{3}\pi L^{*3} \sim B$ for spheres. It follows from equation 7 that the dissymmetry for dilute solutions will also decrease linearly with concentration[25]. For the spherical influenza virus particles in nearly salt-free water L and L^* were calculated[25] from the dissymmetry data[2] to be 128 mμ and 425 mμ respectively. If sufficient electrolyte is added to suppress the electrostatic repulsions, then experimentally it is found that L^* is nearly equal to L. At higher concentrations of particles, corresponding to higher powers in concentration for the osmotic pressure, the dissymmetry increases again as has been observed for nearly salt-free solutions of tobacco mosaic virus[25] and of serum albumin[7]. In the latter case the particles are small ($z \simeq 0$) so that $F^2(z)$ will be nearly unity. However, z^* may be large so that $F(2z^*)$ will be less than unity (for $\theta > 0$).

For any concentration the radial distribution of particles is obtained from a Fourier analysis of the angular scattering data *i.e.* the Fourier inversion of the scattering curve using $F^2(z)$, determined from dilute solutions, as a parameter. The equation of state of the solution *i.e.* the osmotic pressure as a function of concentration, may be expressed in terms of the radial distribution and the interparticle force[26]. No Fourier analysis has yet been made on the scattering data for concentrated solutions but in *Figure 6* is shown the relative scattering calculated for a 60 per cent solution (volume concentration) of spheres[25]. The calculation involves the use of a theoretical radial distribution expected

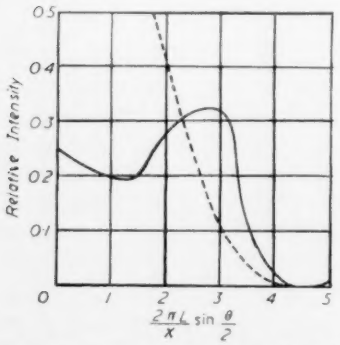


Figure 6. Theoretical scattering per spherical particle for a 60 per cent by volume solution of spheres (continuous line) and for independent spheres (broken line)

for the disordered packing of hard spheres[27]. At a volume concentration of 74·05 per cent the system becomes a cubic or hexagonal system with closest packing of the spheres.

There is a maximum in the scattering curve for $z = 3{\cdot}0$ corresponding, in the form of Bragg's law, to $\lambda' = (1{\cdot}05)2L \sin \theta/2$. The close agreement between Bragg's law and the predicted maximum is coincidental since for repelling particles the scattering curve may show no maximum (for $z > 0$) but only exhibit a shoulder as has been observed for concentrated solutions of haemoglobin[28, 29] and ovalbumin[29]. In *Figure 7* is a small angle scattering x-ray picture taken by RILEY and HERBERT[29] of a 42·4 per cent by weight solution of ovalbumin. The shoulder on the scattering curve occurs at A. Lines at B and C corresponding to Bragg spacings of about 23 Å and 11·5 Å respectively may correspond to prominent interatomic vector distances within the protein molecule. Small angle x-ray scattering curves of red blood cells (haemoglobin content about 33 per cent) also show a shoulder[30] similar to that exhibited by a solution of the same haemoglobin concentration as that in the cells. Some authors[29, 31] have treated the concentrated systems as made up of tiny crystallites with order extending over a few neighbours and applied Bragg's law to the x-ray scattering data. Such a treatment is, however, incompatible with the large intensity of scattering observed at angles near zero. Furthermore, it might be expected that if such crystals were present they would sink to the bottom of the solution since, in general, the molecular packing in the crystallite would be greater than that in the solution. In investigating biological structures, the question of whether true crystallites are present or whether there is random packing of particles can be resolved by performing a Fourier analysis of the scattering data.

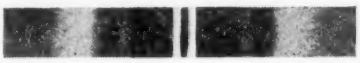


Figure 7. Small angle x-ray scattering for a 42·4 per cent solution of ovalbumin

The problem of evaluating the scattering data for non-independent elongated particles is more difficult than that for spheres. In place of the spherically symmetric radial distribution function one now has two functions, a spatial distribution function and an angular distribution function[25]. The latter defines the fraction of elongated particles having an orientation relative to the orientation of any given particle and may be determined from a fibre-type scattering diagram. BERNAL and FANKUCHEN[32] in their detailed x-ray studies of concentrated solutions (above 13 per cent) of the rod-shaped tobacco mosaic virus imposed a two dimensional crystallinity on the system by orienting the rods by flowing the solution or by inserting it into a fine capillary.

They then found that the particles tend to form a two dimensional close packing array, the interparticle distance being inversely proportional to the square root of the virus concentration. Freshly purified samples of tobacco mosaic virus in water at a concentration of about 2 per cent will form three dimensional crystals in the bottom of the two phase virus system[11, 25]. The interplanar spacings are sufficiently large (340 mμ) to exhibit Bragg diffraction with visible light and the size of the crystals may be determined from the half width of the diffraction peak which is superposed on the scattering curve. Such crystals show spacings similar to those of the three dimensional crystals found in diseased plant cells[33].

REFERENCES

[1] RAYLEIGH, LORD *Phil. Mag.* 41 (1871) 447
[2] OSTER, G. *Chem. Rev.* 43 (1948) 319
[3] — *Science* 103 (1946) 306
[4] PUTZEYS, P. and BROSTEAUX, J. *Trans. Faraday Soc.* 31 (1935) 1314
[5] HELLER, W. and KLEVENS, H. G. *Phys. Rev.* 67 (1945) 61
[6] CAMPBELL, D. H., BLAKER, R. H. and PARDEE, A. B. *J. Amer. chem. Soc.* 70 (1948) 2496
[7] DOTY, P. Unpublished communication
[8] OSTER, G. *J. colloid Sci.* 2 (1947) 291
[9] BIER, M. and NORD, F. F. *Proc. Nat. Acad. Sci. (U.S.)* 35 (1949) 17
[10] EINSTEIN, A. *Ann. Phys.* 33 (1910) 1275
[11] OSTER, G. *J. Gen. Physiol.* 33 (1950) 445
[12] RAYLEIGH, LORD *Proc. roy. Soc.A* 90 (1914) 219
[13] RILEY, D. P. *Brit. sci. News* 3 (1950) 7
[14] SPEISER, R. and BRICE, B. A. *J. opt. Soc. Amer.* 36 (1946) 364
[15] ZIMM, B. H. *J. chem. Phys.* 16 (1948) 1093
[16] OSTER, G. *J. gen. Physiol.* 31 (1947) 89
[17] DEBYE, P. *J. phys. colloid Chem.* 51 (1947) 18
[18] OSTER, G., DOTY, P. M. and ZIMM, B. H. *J. Amer. chem. Soc.* 69 (1947) 1193
[19] JORDAN, W. K. and OSTER, G. *Science* 108 (1948) 188
[20] GUINIER, A. *Thèse Sciences* Paris, 1939
— *Ann. Phys.* 12 (1939) 161
[21] KRATKY, O. *J. Polymer Sci.* 3 (1948) 195
[22] — SEKORA, A. and WEBER, H. H. *Naturwiss.* 31 (1943) 325
[23] GUINIER, A. and FOURNET, G. *C.R. Acad. Sci., Paris* 226 (1948) 656
[24] BERNAL, J. D., FANKUCHEN, I. and PERUTZ, M. *Nature, Lond.* 141 (1938) 523
BOYES-WATSON. J., DAVIDSON, E. and PERUTZ, M. *Proc. roy. Soc.A* 191 (1947) 83
[25] OSTER, G. *International Colloquium on Macromolecules* Amsterdam, 1949
— *Rec. Trav. Chim. Pays-Bas.* 68 (1949) 1123
[26] GREEN, H. S. *Proc. roy. Soc.A* 189 (1947) 103
[27] KIRKWOOD, J. G. and BOGGS, E. M. *J. chem. Phys.* 10 (1942) 394
[28] GUINIER, A. *Bull. Soc. Chim. Belg.* 57 (1948) 90
[29] RILEY, D. P. and HERBERT, D. *Biochim. Biophys. Acta* In press
[30] DERVICHIAN, D. G., FOURNET, G. and GUINIER, A. *C.R. Acad. Sci., Paris* 224 (1947) 1848
[31] KRATKY, O. and POROD, G. *J. colloid Sci.* 4 (1949) 35
[32] BERNAL, J. D. and FANKUCHEN, I. *J. gen. Physiol.* 25 (1941) 111
[33] WILKINS, M. H. F., STOKES, A. R., SEEDS, W. E. and OSTER, G. In press
[34] EDSALL, J. T., EDELHOCH, H. L., LONTI, R. and MORRISON, P. R. *J. Amer. chem. Soc.* In press

4

BIOELECTRIC POTENTIALS, THEIR MAINTENANCE AND FUNCTION

E. E. Crane

A CONSIDERABLE amount of data is available about steady bioelectric potentials, which are the subject of this review, but we still know little about the exact processes by which they are maintained, and even less about the part they play in the functions of the tissues in which they occur. It is comparatively easy to record electric potential differences between two parts of certain organisms, but much less easy to ensure that the potential differences recorded give a reliable estimation of those arising in a specific tissue, and even more difficult, often at present impossible, to interpret them in terms of events taking place within the tissue.

The problem of reliability of potential difference measurements, and of their analysis according to their origins within the tissue, is being solved largely by improvements in technique. The use of simple systems *e.g.* isolated tissue kept under conditions which can be carefully controlled, and under which it functions normally, eliminates a great many uncertainties. Measurements on isolated cells, using micro-techniques recently developed, are sometimes possible and deserve further attention. Another difficult problem in technique which has received some attention and is worthy of more, is the simultaneous determination of electrical properties and the rate at which energy is available to the tissue and the rate at which it is being expended.

The problem of the interpretation cannot be solved by bioelectric measurements alone. Frequently it must await much more information from other fields : cytology, biochemistry, physiology and so on. On the other hand bioelectric measurements can make contributions towards the solution of problems in these fields. Advances in the final understanding of the maintenance and function of potential differences in living tissues are likely to come from a combination of both direct bioelectric measurements and indirect developments in other sciences.

This review is an attempt to summarize some of the available knowledge of steady bioelectric potentials, and to indicate what advances have been made towards the analysis and interpretation of

the ways in which they are maintained, and of the part they play in the functions of the tissues which give rise to them.

HISTORICAL RÉSUMÉ

The foundations of the present subject were laid towards the end of the eighteenth century. There were two crucial experiments, the first of which was performed by WALSH, probably in 1776. In this experiment he succeeded in obtaining a visible spark from the discharge of the *Gymnotus electricus*, thus demonstrating the truth of CAVENDISH's deduction that the shocks produced by certain fish were due to electricity generated by them. No account of the experiment was published by Walsh, but it was described by both HENLY[139] in 1776 and CAVALLO[67] in 1795.

There was a considerable amount of opposition to the view that phenomena could be electrical which originated in so good a conductor as the body of a fish and took place in sea water. COX[77] quotes a contemporary, RONAYNE, who wrote

> Indeed, I must say, that when a gentleman can so far give up his reason as to believe the possibility of an accumulation of electricity among conductors . . ., he need not hesitate a moment to embrace as truths the grossest contradictions that can be laid before him.

The account[6] of the second crucial experiment was published in a supplement to an anonymous tract, the rare *Arco conduttore*, published in Bologna in 1794. The supplement was almost certainly written either by GALVANI or by his nephew ALDINI, or by both of them (FULTON and CUSHING[123], 1936). In Galvani's first two classical experiments in 1791, muscular contractions of a frog were noticed when a copper-iron junction was present. It was VOLTA who rightly identified the junction as the source of the electricity which stimulated the muscle in these experiments, and the experiment described in the anonymous tract[6] was probably carried out in reply to VOLTA's criticism[271] (1792). It was important because it demonstrated that electricity produced in the frog itself could cause the contraction of frog muscle—and this was the most sensitive method available at the time for detecting electric current.

The subsequent development of VOLTA's ' organe electrique artificiel' *i.e.* the Voltaic pile[272], was probably based on his study of the ' organe electrique naturel ' of the torpedo, the laminar nature of which HUNTER[159] (1775) had already described in detail (WALKER[274]).

Apart from its influence on the development of experimental physics, the study of electric fish had an important influence on the development of theoretical physics. CAVENDISH, who as far as we know never handled or even saw a live torpedo, carried out a number

of experiments on models in air and salt water, as a result of which he made clear the distinction between potential difference and quantity of electricity, and moreover, years before the discovery of the Voltaic pile or thermoelectricity, anticipated the idea of electromotive force (see CAVENDISH[68] ; MAXWELL[201] ; COX[77]).

During the nineteenth century the study of the production of electric currents by any living organism held great fascination for physicists and biologists alike. FARADAY[111] wrote in 1839

> Wonderful as are the laws and phenomena of electricity when made evident to us in inorganic or dead matter, their interest can bear scarcely any comparison with that which attaches to the same force when connected with the nervous system and with life.

The investigations carried out were descriptive rather than quantitative, and most workers were content to record the existence and approximate magnitude of the potential difference between two particular regions of a living system, sometimes under various conditions. On the other hand some shrewd guesses were made as to the possible function of bioelectric currents ; for instance DONNÉ[102] (1834) wrote ' sont-ils capables de rapprocher, de séparer les molécules l'une de l'autre, et de donner naissance à des corps nouveaux, comme cela arrive dans la règne inorganique ? ' DU BOIS-REYMOND (1848) made valuable contributions to various branches of the subject[103]. Much of the later experimental work has been reviewed by BIEDERMANN[23] (1895), and WALLER[275] (1897) has given an account of the ' bitter but fruitful controversies ' which raged as to the source of the currents observed.

During the latter part of the nineteenth century scientists were much occupied with the measurement of action and injury currents of nerve and muscle, and the increasing sensitivity and decreasing time of response of instruments for the measurement of voltage and current were determining factors in the development of the study of bioelectricity (PLEDGE[229], 1939 ; MATTHEWS[200], 1948). Instruments which could be used to measure the small potential differences produced by living tissues were developed, and it was no longer necessary to use frog muscle as a detector. The development of the subject largely followed the improvements in technique. From 1820 onwards galvanometers were available of progressively greater sensitivity. Later, instruments based on LIPPMANN's[188] (1875) capillary electrometer, and at the turn of the century the string galvanometers of EINTHOVEN and others, made it possible to measure fractions of a millivolt (see also ETTISCH and PÉTERFI[106]).

The use of valve amplifiers from 1920 onwards provided a great impetus to the study of bioelectric phenomena, for instruments could be chosen for speed of deflection and convenience, as sensitivity became

a matter of secondary importance. Microvolts could be measured, and the additional use of a cathode-ray oscillograph reduced the time of response to 10^{-4} sec or less ; MATTHEWS[200] gives a general discussion. The limit of sensitivity is now not instrumental, but is imposed by spontaneous electrical fluctuations of the same character as Brownian movement[211].

OCCURRENCE OF BIOELECTRIC POTENTIALS

The present review is concerned with the steady electric potential differences which are maintained in living tissues. Rapidly fluctuating potential differences also exist and are of great importance, but except for comparisons of orders of magnitude they will not be discussed here. These periodic potential differences are usually classified according to the organ which gives rise to them, and they fall into fairly well defined subjects, each with an extensive literature of its own. The main subdivisions are given in *Table I.*

Table I. Approximate Magnitude and Frequency of Important Classes of Bioelectric Potentials

Description (or name of subject concerned)	*Source*	*Order of magnitude mv*	*Frequency cycles/sec*
Steady potential differences	*Various plant and animal tissues*	*up to ca* 150	0
Electrocardiography	*heart muscle*	*ca* 30 (*ca* 1 *from external leads*)	1 *to* 3
Electroencephalography	*brain*	*ca* 0·05 *from external leads*	0 *to* 60 (10 *dominant*)
Nerve impulses	*nerve*	*up to ca* 100	*ca* 100 (*up to* 1,000, *and up to* 20,000 *for mammalian ear*)
Action potentials in skeletal muscles		*up to ca* 100	5 *to* 200 +
Discharges from electric organs of some fishes and elasmobranchs		*up to ca* 600,000 (*up to ca* 150 *from each electroplax*)	*discharge frequency ca* 200

In *Table I* no separate entry is made for injury potentials, most of which come under the first entry. They are produced by almost any type of injury, but only in living systems. The subject has been reviewed recently by HEILBRUNN[138] who gives a list of injury potentials in plants ; they are mostly between 30 and 100 mv.

The term action potential is often used loosely to cover any potential difference, whether periodic or aperiodic, which results from stimulation. There are for instance aperiodic variations of the potential

difference measured from the skin surface (*e.g.* psycho-galvanic reflex) following emotional or other suitable stimulation. These potential responses, which are complex in character and may last for well over a minute, have amplitudes up to about 10 mv.

The primary interest of the resting potentials of nerve and muscle has been in relation to their action potentials, and their detailed consideration would extend the scope of this paper too far. Only a few aspects of problems concerning them are included. The subject has been reviewed recently by GRUNDFEST[133].

The potential differences exhibited by the electric organs of some fishes and elasmobranchs are listed separately in *Table I* although the potential differences produced are probably relatively constant even during discharge (see pp 115, 116). The high electromotive forces are produced by a large number of units (electroplaxes) in series ; each electroplax does not contribute more than about 150 mv. It is clear therefore from *Table I* that the electromotive force generated by a single biological membrane does not greatly exceed 100 mv.

In comparison with the compactness of most of the subjects represented in *Table I*, the study of problems associated with steady potential differences is less well coordinated. There are several reasons for this : they have various underlying causes still to a certain extent unelucidated, they occur in a very large number of living tissues and are in some cases critically dependent upon experimental conditions, they are often difficult to analyse and their sources difficult to identify. It is in many ways easier to measure rapid changes of potential differences than to measure the steady potential differences on which the changes are superimposed. To workers investigating these changes, the underlying ' steady ' potential differences (although of considerable interest in their own right) merely form a background, undergoing troublesome slow changes which must be eliminated in measurements.

Definitions— Alternatives to the name steady potentials are resting potentials used *e.g.* by CURTIS and COLE to differentiate them from action potentials, base-line potentials used by MEADER and MARSHALL and others, direct potentials (in distinction to alternating potentials) and standing potentials. The first name is perhaps preferable to any of the others, in spite of the fact that the potentials do undergo slow temporal changes ; the other names were all chosen for a secondary reason.

The term potential itself has of course no meaning unless a statement is given of the zero with respect to which the potential is measured, and for this reason the full expression ' potential difference (pd) between A and B ' should be used. There are, however, occasions

when it is convenient to regard the electrode at one point, say A, as at a constant potential (this is sometimes called the indifferent electrode) and, in order to avoid prolixity, to refer to the potential difference between the electrodes at B and A as the potential at B. The abbreviation implied should however be constantly borne in mind (see also p 99).

It is important to realize that potential differences are normally measured directly only between two metallic conductors. It is necessary to make electrical contact with biological systems by means of electrodes, and in general the pd measured will depend upon the type of electrode and on the characteristics of the system, and results with one arrangement are not immediately comparable with those obtained with another (see p 98). Electrodes used for the measurement of steady potential differences in biological systems are usually calomel or Ag/AgCl, but others have been used for special purposes.

The use of the term bioelectric current for bioelectric potential, which is still common with some authors (see ROSENE[250]), is wrong and may be misleading. The current produced in an external circuit by a tissue depends not only on the pd it would exhibit on external open circuit, but also on the values of the resistances involved. The use of the term bioelectric current for bioelectric potential is a legacy from the time when bioelectric effects were studied by current measurements ; nothing is gained by its continued use.

In this review the sign of a potential difference is regarded from the point of view of the external circuit : ' the potential at B is positive with respect to that at A ' means that if A and B were connected by a circuit external to the organism in which they are situated, a positive current would flow through this external current from B to A.

A statement of policy must be made on one other problem in terminology : electromotive force (emf) and potential difference (pd). The emf produces a difference of potential between any two points in the circuit which depends upon the distribution of resistance in the circuit. In an open circuit the total resistance is infinite, no current is maintained by the emf and the pd measured is the total emf. Usually however this limiting condition is not realized in biological material, often because local short circuits exist, and the pd measured does not represent the whole of the emf generated. It seems more correct therefore to avoid the term emf even in cases where no external current is drawn from the tissue, unless it is known that internal currents are also negligible.

Steady bioelectric potentials from various sources— The magnitude of the steady potential differences from a number of plant and animal tissues are given in *Table II.* Here and elsewhere names of animals are taken

Table II. Sources of Bioelectric Potentials

Species	*Structure*	*Positive side*	*Negative side*	*Approximate pd (mv)*	*Temperature °C*	*Reference*	*Approximate Q_{O_2} (μl/mg dry wt/hr)*	*Temperature °C*	*Reference*
Valonia macrophysa	*cell wall*	*external sea water*	*protoplasm*	20–50		264			
		interior sap	*external sap*	62		93			
		interior sap	*external sea water*	8		93			
Halicystis ovalis	*cell wall*	*external seawater*	*interior sap*	60–80		27			
Chara *sp.*	*cell wall*	*external tap-water*	*interior (injury?)*	140		184			
Onion (Allium cepa)	*root (apical 20 mm)*	*apex*	*base*	3 *per mm*		194			
	root (apical 6 mm)	*apex*	*base*	10 *per mm*		194			
Various plants	*root*						— 1 *to* — 3		179
Slime fungus	*plasmodium*	*interior*	*exterior*	2		184			
Amoeba proteus	*cell membrane*	*inside or outside*	*inside or outside*	1–3		40	— 0·8	20°	104
Amoeba sphaero-nucleolus	*cell membrane*	*inside or outside*	*inside or outside*	0·5–1·5		40			
hydroid Tubularia	*stem*	*rest of stem*	*hydranth*	1		16			
Bivalves:									
Anadonta *sp.* Unio *sp.*	*mantle*	*shell-secreting surface*	*chamber surface*	5–40		192			
Ostrea edulis	*mantle*						— 0·96, — 1·30	27·5°	70
Frog (Rana temporaria)	*isolated skin*	*serosa*	*epidermis*	10–80		118	*dorsal* —0·87 *ventral* —0·97 *leg* —1·15	25°	230 *see also* 156
Rat (Wistar)	*skin, wounded (superficial cut)*	*wound*	*neighbouring skin*	10–30		65			
Guinea-pig	*skin, wounded (incision traversing skin and subcutaneous fascia)*	*wound*	*neighbouring skin*	*up to* 10		51			

Table II. Sources of Bioelectric Potentials (continued)

Species	*Structure*	*Positive side*	*Negative side*	*Approximate pd (mv)*	*Temperature °C*	*Reference*	*Approximate Q_{O_2} (µl/mg dry wt/hr)*	*Temperature °C*	*Reference*
Human	*epidermis (sliced)*						−1·0 *to* −3·8	37·5°	279
							− 1·5	37·0°	41
Human	*skin*	*measured between various external points*		*up to* 30 *or* 80		115 255			
Human	*skin, wounded (deep incision)*	*skin close to wound*	*skin distant from wound*	3		84			
Human	*finger skin, wounded (sandpaper abrasions)*	*wounded finger*	*corresponding finger of other hand*	15–50					10
Frog (Rana temporaria)	*isolated oesophageal mucosa*	*submucosa*	*mucosa*	2	25°	87	− 0·8	25°	96
Frog (R. temporaria) / *Toad* (Bufo bufo)	*isolated gastric mucosa*	*submucosa*	*mucosa*	20 *to* 40	25°	89, 260	− 2·1	25°	95
Polecat (Mustela putorius)	*isolated gastric mucosa*	*submucosa*	*mucosa*	10 *to* 30	38°	87	− 8·2	38°	95
Dog	*stomach*	*serosa*	*mucosa*	70 *to* 95	38°	238, 243			
Frog (R. temporaria)	*isolated pyloric mucosa*	*serosa*	*mucosa*	5	25°	87	− 1·2	25°	94
Frog (R. temporaria)	*isolated duodenal mucosa*	*submucosa*	*mucosa*	2	25°	87			
Frog (R. temporaria)	*isolated rectal mucosa*	*serosa*	*mucosa*	10 *to* 30	25°	87	− 3·1	25°	96
Frog (R. temporaria)	*isolated muscle layer*			0	25°	89	− 0·6	25°	96
Guinea-pig	*isolated atrium*			0	38°	87			
Rat	*isolated diaphragm*			0	38°	87	− 6·8	38°	204
Rat	*isolated kidney capsule*			0	25°	87			
Goat	*isolated kidney capsule*						− 1·0	38°	177

Goat	*isolated kidney capsule*						− 1·0	38°	177
Perch (Perca fluviatilis)	*isolated swim bladder*			0	25°	87	− 0·85	25°	96
Chick	*whole embryo* (0→40 *somites*)	*cephalic end of nervous system*	*caudal end*	0·01→1		54			
	whole embryo :								
	0·085 *to* 1·86 *mg dry wt*						− 14·3 *to* − 22·4	37·5°	182
	0·4 *to* 4·7 *mg dry wt*						− 8·1 *to* − 21·9	37·5°	183
Pig	*choroid plexus*	*epithelium*	*stroma*	230*		258			
Rabbit	*choroid plexus*						− 20	37·5°	180
Rabbit	*ciliary body*	*epithelium*	*stroma*	230*		120			
Squid (Loligo pealii)	*nerve fibre (giant axon)*	*external sea water*	*internal axoplasm*	41		145			
				51		92			
Squid (Loligo forbesi)	*nerve fibre (giant axon)*	*external sea water*	*internal axoplasm*	48	18–23°	148			
Electric eel (Electrophorus electricus)	*main electric organ*	*anterior end*	*posterior end*	300 *to* 600 × 10^3 (140 *across one electroplax*)		72, 69	− 0·3	29·5°	69
							− 0·65		
							− 0·3	*not given*	217
	bundles of Sachs	*anterior end*	*posterior end*	40 × 10^3 (100 *across one electroplax*)		72	− 0·3	*not given*	217
Electric catfish (Malapterurus electricus)	*electric organ*	*posterior end*	*anterior end*	300 *to* 400 × 10^3		19			
Torpedo occidentalis	*electric organ*	*dorsal surface*	*ventral surface*	200 × 10^3		80			
Narcine brasiliensis (*Olfers*)	*electric organ*	*dorsal surface*	*ventral surface*	20 × 10^3 (100 *across one electroplax*)		78			

* Difference in oxidation-reduction potential, estimated with dyestuffs

from the British Museum *List of British Vertebrates*[38] (1935). Rates of oxygen uptake are also given, expressed in terms of the Q_{O_2} (μl oxygen output per mg dry wt per hr). *Table II* does not contain details of all tissues on which measurements have been made, nor are references included to all papers describing work on the tissues listed. It is intended as a summary of some of the information on tissues discussed in this paper and on a few others given for comparison. Further references may be obtained by consulting ROSENE's bibliography of continuous bioelectric currents and bioelectric fields in animals and plants[250] (1947), which covers more or less the same ground as this review. It contains references and titles of 1,400 papers and books in alphabetical order of author. It is indexed, but in general only from words occurring in the titles of the publications, and not from their content.

Results entered in *Table II* were obtained from measurements carried out under conditions assumed to be optimal, or normal. There is a wide variation of pd and Q_{O_2} with temperature and other factors, and in some cases the pd varies greatly with the ionic composition of the solutions bathing the tissue ; figures in *Table II* can therefore only serve as guides, and the original papers should be consulted for details of experimental technique. Where a range is given it represents not extreme values, but those commonly occurring.

The direction of the pd is in many cases the same in different examples of similar tissue. For instance in all electric rays in which the electric organ is developed from the jaw muscles, the dorsal surface is positive and the ventral surface negative. Throughout the gastro-intestinal tract of the frog the nutrient side is positive with respect to the outer surface of the mucosa, and in frog skin also, the side supplied with blood vessels is the positive one. An injured portion of nerve or muscle is negative with respect to a neighbouring uninjured portion ; a wound in the skin of such animals as have been studied is however positive with respect to the neighbouring intact skin until healing is well advanced (see p 109). The pd across the walls of large plant cells may be in either direction according to species and the conditions under which the determination is made.

Table II shows that the steady pd developed across active living membranes is commonly of the order of tens of millivolts and is in some cases even more than a hundred millivolts. The entries show that there is by no means a simple relationship between the rate at which oxygen is utilized and the pd developed ; energy is expended in various ways, not all of which involve the maintenance of a steady pd. This problem is considered further in a later section.

The functions of the active structures represented in *Table II* are

various. In many cases the function is secretion and/or absorption ; these are discussed later (pp 117-123). There is however no question of electrical activity being related only to a single function. One of the most important problems concerning bioelectric potentials is their part in various biochemical processes. This problem will be discussed for structures for which some evidence is available, but it is as yet largely unanswered. It offers a profitable field for study, if suitable experimental techniques are used, and one which has been very little explored.

CAUSES OF BIOLOGICAL POTENTIAL DIFFERENCES

Our understanding of the causes underlying the existence of potential differences in living organisms is still very incomplete. We know that certain asymmetries in structure, composition and permeability, which are likely to be sources of electromotive forces, do occur in living systems, and that the emf's should be capable of producing external potential differences. Moreover, it is often possible to build inanimate models which develop emf's of similar orders of magnitude and modes of variation (*e.g.* BEUTNER[20] ; MICHAELIS[205, 206]). This is, however, not the whole story. The fact that a model has some of the properties of the living system is not proof that their modes of action are identical, and extreme caution should be applied in using the behaviour of models as evidence for the existence of similar conditions in the living structures.

It is certain that potential differences observed in living tissues have no single simple origin. Even those actually produced within a tissue, and by no means all experiments have been successfully designed to measure only these (see OGSTON[222]), may arise from a variety of conditions. The only factor common to all such conditions is their asymmetry, but this may be an asymmetry of structure, permeability, ionic concentration or function.

Asymmetrical conditions which are known to be a source of emf and which occur in living organisms include :

1. a liquid junction between two solutions of different composition and/or concentration
2. a membrane separating solutions of different composition and/or concentration
3. a phase boundary *e.g.* between an electrolyte solution and a liquid solvent immiscible with water
4. an oxidation-reduction system
5. the relative motion of a solid and liquid.

A general quantitative discussion of the potentials produced under these conditions and their applications to biological systems is given

in books on physical biochemistry (BULL[42], 1943 ; HÖBER[142], 1946). Certain points of importance and of recent interest are considered below.

Liquid-junction potentials (*diffusion potentials*)— A pd is developed at the junction of two aqueous solutions of different composition and/or concentration, unless all the ions present have the same mobility. Mobilities of ions most important in biology are given in *Table III*.

In the simplest possible case, that of a liquid junction of two solutions of the same univalent salt, whose concentrations (or more correctly activities) are C_1 and C_2, the diffusion potential E is given by:

$$E = \frac{RT}{F}\frac{u^+ - u^-}{u^+ + u^-}\ln\frac{C_1}{C_2} \qquad \ldots.(1)$$

where R = gas constant = 8·31 joule/mol/°C
T = absolute temperature = 298° for 25°C
F = Faraday constant = 9·65 × 10^4 coulomb/mol
u^+ = mobility of cation
u^- = mobility of anion

Thus at 25°C,

$$E = 59\left(\frac{u^+ - u^-}{u^+ + u^-}\log_{10}\frac{C_1}{C_2}\right) \text{ mv.}$$

Table III. Mobilities of Ions at 18°C, and Temperature Coefficients of Mobilities

Anions	*Mobility*	*Temp. coeff. per °C*	*Cations*	*Mobility*	*Temp. coeff. per °C*
OH^-	174	·018	H^+	315	·016
Cl^-	65	·022	K^+	64	·022
I^-	66	·021	NH_4^+	64	·022
HCO_3^-	41	·022	Na^+	43	·024
NO_3^-	62	·021	$\frac{1}{2}Ca^{++}$	51	·025
$C_2H_3O_2^-$	35	·024			
$\frac{1}{2}CO_3^{--}$	60	·022			
$\frac{1}{2}SO_4^{--}$	68	·023			

(The mobility of an ion is its equivalent conductance at infinite dilution *i.e.* conductance in ohm^{-1} of solution containing 1g equivalent of the ion between two electrodes 1 cm apart.)

Table III shows that in KCl and NH_4NO_3 the cation and anion have nearly equal mobilities. Since NH_4NO_3 is more reactive than KCl, a concentrated (usually saturated) solution of KCl is generally used for bridges connecting half-cells to the solutions bathing the two sides of a biological membrane. Potential differences produced at the liquid junctions in the circuit are thus reduced to a minimum. The diffusion potential between two solutions of KCl, one ten times as concentrated as the other, is only 0·4 mv; NaCl solutions with the same concentration ratio produce 3·2 mv.

When a membrane is injured, solutions normally separated are brought into contact and one might expect to observe diffusion potentials originating at the liquid junctions created. The application

of diffusion theory has met with some success in accounting for certain injury potentials. For instance STEINBACH[257] (1935) measured the variation of injury potentials of the scallop adductor muscle in various electrolytes at different concentrations. He found that his experimental results were consistent with the mode of variation (for both electrolyte and its concentration) predicted from equation 1, but he found it necessary to assume the presence in the tissue of an anion, possibly a protein, with a low mobility.

It is clear, however, that simple diffusion theory cannot account for most bioelectric potentials. DEAN and GATTY[99] (1937) discussed its inadequacy with special reference to frog skin, and gave an extensive bibliography.

Membrane potentials— DONNAN[101] laid the foundation of the study of membrane equilibrium forty years ago, and certain biological applications of his theory have proved fruitful *e.g.* the ionic exchange in red cells during the respiratory cycle (VAN SLYKE[267], 1926 ; see also LOEB[189], 1922). The following is a simple example of a Donnan-equilibrium membrane potential :

If a membrane is permeable to Na^+ and Cl^- ions but not to P^- (say a protein) and is initially bathed on one side *a* by a solution containing Na^+ and Cl^-, and on the other *b* by a solution containing Na^+ and P^- then, when Donnan equilibrium is attained, a pd *E* can be measured between calomel electrodes connected by saturated KCl bridges to the two solutions where

$$E = \frac{RT}{F} \ln \frac{[Cl^-]_b}{[Cl^-]_a} = 59 \left(\log_{10} \frac{[Cl^-]_b}{[Cl^-]_a} \right) \text{ mv} \qquad \ldots . (2)$$

$$\text{or} \quad 59 \left(\log_{10} \frac{[Na^+]_a}{[Na^+]_b} \right) \text{mv.}$$

R, *T* and *F* are defined on p 96 and square brackets denote concentrations, or more correctly activities.

Those who hoped that the theory could be used to account for all bioelectric phenomena have been disappointed. Donnan membrane potentials cannot provide a continuous supply of electrical energy and for this reason are unlikely to be important for one of the most significant classes of bioelectric potentials, as will be shown later.

The principles of Donnan membranes do not apply to living cells except in special cases, and many complications arise. For instance electrodes sensitive to any ion to which the membrane is permeable cannot record the corresponding Donnan membrane potential (BULL[42], 1943). This fact has been used by DAVIES and OGSTON[97] (1950) to obtain information about a system containing a selectively permeable membrane, by the use of simple reversible electrodes. They measured

the steady pd across isolated frog gastric mucosa (not secreting acid) with *a* calomel electrodes, connected to the solutions bathing the mucosa by saturated KCl bridges, *b* Ag/AgCl electrodes and *c* high-conductivity glass electrodes ; *b* and *c* were placed directly in the solutions bathing the mucosa. The experimental results given in *Table IV* refer to the constant and reproducible pd between the secretory and nutrient solutions at 25°C. These were 0·012 M NaCl made up to 0·12 M with $NaNO_3$ and at *p*H 5·60 on the secretory side, and bicarbonate saline (0·095 M chloride), *p*H 7·40, on the nutrient side. The last column shows that the differences are entirely accounted for by the difference in the concentrations of chloride ions ($b - a$) and of hydrogen ions ($c - a$) in the two solutions.

Table IV. Potentials across Frog Gastric Mucosa

	Electrodes	*Observed pd mv*	*Difference calculated from equation 2 mv*
a	*calomel*	−22	
b	Ag/AgCl	+30	$b - a = +52$
c	*high conductivity glass*	+84	$c - a = +106$

The general conditions which govern the potential difference between a pair of such electrodes are as follows. If two phases, separated by a membrane, are at equilibrium with respect to all those species of ions (A, B, C...) to which the membrane is permeable, the pd observable under different conditions between two electrodes one in each phase is given in *Table V*.

Table V. General Conditions for pd in Selectively Permeable Membrane System

Membrane permeable to	*Membrane impermeable to*	*Electrodes reversible to*	*Pd observable*
none	A, B, C.....	*any*	—
A	B, C.....	A	*zero*
A	B, C.....	*any except* A (*say* B)	*finite and reproducible*
A	B, C....	*any except* A (*say* B)	*zero**

**Special case where* B *is distributed as it would be at equilibrium were the membrane permeable to it.*

The experiment just described therefore shows that isolated frog gastric mucosa not secreting acid is effectively impermeable to both Cl^- and H^+ ions.

If the two phases are not at equilibrium with respect to A, B, C..... and if passage of a current between the electrodes facilitates the transfer of ions with respect to which the disequilibrium exists, then a reproducible potential difference is observed. The magnitude of the pd depends upon the electrodes, and can be calculated only in simple cases. If the transport number of the ion to which the electrodes are reversible is zero, this case is equivalent to the fourth given in *Table V*.

BERNSTEIN[17] (1912) first developed the membrane theory of bio-

electric potentials in relation to general conditions, including those producing action currents. He regarded membranes in living tissue as permeable to cations (especially K^+) but not to anions, and their action currents as due to relative changes in permeability. A difference in permeability to positive and negative ions results in the formation of a double layer of changes of opposite sign, separated by the membrane, and hence of a potential difference across it. MICHAELIS[205] developed the idea of the differential permeability of a membrane as due to the charges in the pores, a positive charge favouring the transmission of anions and a negative charge favouring cations, and he has shown that this effect is possible with collodion films.

Much experimental evidence has been produced (see HEILBRUNN[138], 1947 for summary) both for Bernstein's development of the membrane theory, and against it. OGSTON[222] goes so far as to object to the whole concept of a membrane potential, on the grounds that a single electrode potential is a meaningless term, any pd being necessarily measured between two electrodes.

Methods for the potentiometric analysis of membranes in terms of layers of different selective permeabilities have been developed by OSTERHOUT[224] (see bibliography by ROSENE[250]), and MEYER[202], MEYER and BERNFELD[203], USSING[265, 266]. Much information has been obtained by these and other workers about the effects of changes in the ionic composition of the surrounding fluids on the pd across membranes, and some of this information has been used to compare the selective permeability of different layers of the membrane and to locate the boundary at which the pd is developed (see also HÖBER[142]).

KATZ[163] (1948) has discussed the evidence which serves to identify the membrane or interface, at which the potential changes in muscle occur, with the surface of the muscle fibres. The evidence comes from his own experiments and from others such as those by GRAHAM and GERARD[130] (1946) who made direct measurements of resting and action potentials with a microelectrode inserted in a muscle fibre.

In a recent paper of some importance HODGKIN and KATZ[148] showed that the observed reversal of membrane potential during the active phase of the squid giant axon could be quantitatively explained on the assumption that the resting membrane is more permeable to K^+ than to Na^+, and the active membrane more permeable to Na^+ than to K^+. They suggested that the Na^+ permeability is greatly increased during the active state and that the K^+ permeability is virtually the same in both active and resting states. Radioactive tracer studies have confirmed the change in Na^+ permeability (KEYNES[166]) but have shown (KEYNES[165]) that K^+ permeability also changes.

The report of the discussion on the properties and functions of

natural and artificial membranes, held by the Faraday Society in 1937, forms a valuable contribution to the subject of membrane potentials. In the opening paper KROGH[181] (1937) emphasized the distinction between membrane function which passively allows some substances through and holds others back, and active transport which requires the expenditure of energy (see also KROGH[182], 1946). Energy relations, and the relationship between active transport and potential differences, will be discussed later.

Phase-boundary potentials— Many phase boundaries (interfaces) occur in living tissue, and some are the source of potential differences. The idea that the interfaces between thin layers of oil-like solvents and aqueous solutions of different concentrations are the source of bioelectric potentials has been put forward by BEUTNER[20] (1920) and others as an alternative to the membrane theory. BEUTNER[20-22] has carried out a great deal of work on living tissues and on models with various oil/aqueous solution interfaces, which he considers to behave similarly. OSTERHOUT[225, 226] has worked with models devised to represent giant plant cells.

BARNES and BEUTNER[12] (1946) have shown that acetylcholine has a very marked effect on the potential difference developed at an oil/saline interface. They have suggested that this effect is the basis of nerve action potentials, and that these are phase-boundary potentials produced by acetylcholine in contact with nerve lipoid. While the experimental results obtained by these workers are undoubtedly of importance and have aroused considerable interest, the authors' exclusive interpretation of them (' the old Bernstein concept of an ionic sieve membrane is untenable on both theoretical and experimental grounds ') has not been universally accepted. There is much to be said on both sides, and it seems likely that at the moment we know too little of the fundamental reactions to insist on the acceptance of any one universal theory.

OGSTON[222] (1947) has pointed out that particular difficulty is attached to the interpretation of pd's measured with electrodes in different phases, in terms of the pd's originating in the interface itself and those originating outside it. He also criticized the method of CRAXFORD, GATTY and TEORELL[91] (1938) of placing the electrodes in the ' interphase '.

Oxidation-reduction potentials— A pd is observed when an inert electrode uch as gold or platinum, immersed in a reversible oxidation-reduction system such as a solution of ferrous and ferric chlorides, is connected to a normal hydrogen electrode. This pd is referred to as the oxidation-reduction potential of the system, the potential of the normal hydrogen electrode being defined as zero.

At a fixed *p*H the observed potential E_h of a reversible oxidation-reduction system is given by

$$E_h = E'_0 + \frac{RT}{nF} \ln \frac{[\text{ox.}]}{[\text{red.}]} \qquad \ldots . (3)$$

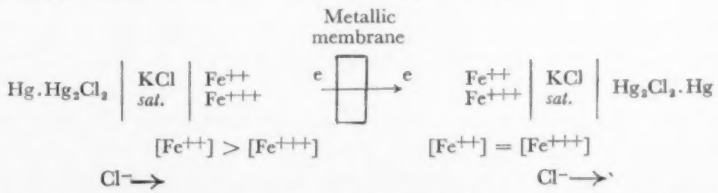
where R, T and F are as defined on p 96 ; n is the number of electrons involved in the oxidation *e.g.* $n = 1$ for the (Fe^{++}, Fe^{+++}) system ; square brackets indicate concentrations (correctly activities) ; E'_0 is the potential at *p*H 7·0 and for unit activities, when [ox.] = [red.]. *Table VI* gives values of E'_0 for some biologically important systems.

Table VI. Oxidation-Reduction Potentials of Some Biologically Important Systems

	System	E'_0 *mv*	T °C
H_2O	$\frac{1}{2}O_2 + 2H^+ + 2e$	815	25
Ferro oxidase	*Ferri oxidase* + e	?	
Ferro cytochrome a	*Ferri cytochrome a* + e	290 (?)	25
Ferro cytochrome c	*Ferri cytochrome c* + e	260	25
Ferro cytochrome b	*Ferri cytochrome b* + e	− 40 (?)	25
Succinate	*Fumarate* + $2H^+$ + 2e	− 26	25
Reduced flavin enzyme	*Flavin enzyme*	− 63	38
Malate	*Oxalacetate* + $2H^+$ + 2e	−102	37
Lactate	*Pyruvate*	−180	35
Alcohol	*Acetaldehyde* + $2H^+$ + 2e	−190	30
β-Hydroxybutyrate	*Acetoacetate* + $2H^+$ + 2e	−293	38
Reduced coenzyme I	*Coenzyme I* + $2H^+$ + 2e	−290	30
H_2	$2H^+ + 2e$	−414	25
Formate	$CO_2 + H_2$	−420	38

Quoted from GODDARD[133], based on data from BARRON[13, 14] and GREEN[131]

As recently as 1946, HITCHCOCK[141] stated 'In the absence of a metal electrode bathed by a solution containing the oxidant and reductant in question, there is no mechanism for the existence of an oxidation-reduction potential ; accordingly bioelectric potentials cannot be oxidation-reduction potentials'. However KORR[175] pointed out in 1939 that the oxidation-reduction potential of the following system can be measured with non-metallic *i.e.* calomel, electrodes



A similar argument applies to a biological system provided with an electron-transfer mechanism which could replace the metallic membrane, so that oxidation-reduction potentials could be manifested as bioelectric potentials. LUND[190] (1928) has for long held that this is so. However, oxidation-reduction potentials must not be identified with potential differences which could be measured between calomel

electrodes, without supporting evidence. UMRATH[263] (1933) and BLINKS and PICKETT[31] (1940) investigated the relationship between oxidation-reduction potentials in Nitella and those measured with calomel electrodes and KCl bridges ; they showed that one could remain constant during considerable changes in the other. STIEHLER and FLEXNER[258] (1938) and FLEXNER and STIEHLER[114] (1938) used indicators to measure oxidation-reduction potentials in the epithelium and the stroma of the choroid plexus of the pig, and found the epithelium to be 230 mv more positive than the stroma. Similar results were obtained by FRIEDENWALD and STIEHLER[120] (1938) for the ciliary body of the eye of the rabbit. No conclusions can be drawn from these experiments alone about the pd observable between calomel electrodes connected to solutions bathing the epithelium and the stroma, since the relevant electrical characteristics of the intervening structure are unknown.

Recently JOSEPH *et alii*[162, 161] (1948, 1949) have determined the pd across the synovial membrane in the knee joint of the dog using saturated KCl-calomel electrodes, in the presence of a wide range of metabolic inhibitors. They found that, at concentrations many times greater than those producing almost complete inhibition of cellular respiration, pd's up to 400 mv were recorded across the membrane, and they relate these pd's in some detail to the energy levels in the cytochrome system. It is very important that these experiments should be repeated with other membranes.

Electrokinetic potentials— These potentials, also known as streaming potentials, are set up between the ends of a capillary tube when liquid moves through it. Their production is the opposite of electro-osmosis, or electroendosmosis, which is the motion of liquid through a capillary tube whose ends are maintained at a difference of potential. Both are due to the existence of electrically charged layers of opposite sign at the boundary between the liquid and the walls of the tube. Streaming potentials are set up by the relative displacement of these charged layers ; electro-osmosis takes place by virtue of their displacement caused by the applied pd.

According to simple theory the potential difference V between the ends of a single tube is given by

$$V = \frac{\zeta K p}{4\pi \nu K} \qquad \ldots\ldots(4)$$

where K = dielectric constant
ν = viscosity
K = specific conductance
} of the solution
p = pressure difference between the ends of the tube
ζ = zeta potential.

The zeta potential is the pd between the solution and the wall of the capillary tube. It is zero at the iso-electric point, and lies between about + 50 mv and − 50 mv for many interfaces at other values of the *p*H. According to simple theory V is independent of the length, cross section and shape of the capillary tube.

Similar results may be obtained for a diaphragm containing a number of pores (see GLASSTONE[124] (1940) for further discussion of the theory and non-biological applications).

Many difficulties appear in the application of electrokinetic theory to observed phenomena, especially in the biological field (BULL[42], HITCHCOCK[141]). One complication is the 'important and somewhat disconcerting conclusion' reached by BULL that the classical theory, including equation 4, does not apply to very small capillaries. If the radius of the capillary is less than five times the thickness of the electrical double layer, serious deviations from the simple theory occur, and the true zeta potential is much smaller than that calculated from equation 4 (BULL and MOYER[43], 1936; KOMAGATA[174], 1934).

Streaming potentials of a few millivolts are produced by the movement of blood, and MILLER and DENT[207] (1943) have claimed that they are responsible for the T-wave in an electrocardiogram.

Electro-osmosis and electrophoresis do not come within the frame of reference of the present review; reference may be made to ABRAMSON[1] (1934) and ABRAMSON, MOYER and GORIN[2] (1942). MUDD[212–214] studied electro-osmosis through mammalian serous membranes, but he did not pursue theoretical interpretations very far. He did not explicitly state that his results show that 'action currents' of glands influence the process of secretion, but he went so far as to say that their consideration 'endows this possibility with a degree of probability amounting almost to certainty'.

General considerations— The origin of each of the five classes of potential difference discussed above (except the last) has been held to be the general, or at least the most important, source of bioelectric potentials by one or more groups of workers. The publicity each theory has received has depended almost as much on the strength of personality of its protagonists as on its intrinsic merit.

In a subject such as the present one, which is interlocked with many other divergent subjects, there is inevitably a tendency to seek to explain phenomena in terms of current developments of interrelated subjects. It seems fairly certain that potential differences are produced in living organisms by all the classes of asymmetry discussed above, or at any rate by asymmetries currently described in those terms.

It seems likely that some of the most useful lines of attack on the problems of the mechanisms of the production and maintenance of

bioelectric potentials are at present indirect ones. A greater understanding of bioelectric phenomena will surely come through knowledge of the molecular, atomic, ionic and electronic reactions which occur in the tissues giving rise to them, and of the energy exchanges in these reactions. This knowledge is likely to be obtained by approaches from many different angles, including experiments with microtechniques on isolated tissue and isolated cells.

BIOELECTRIC POTENTIALS AND PHYSIOLOGICAL CHANGES

This section contains a brief account of recent work on the correlation between changes in steady potential differences and concomitant physiological changes. Progress has certainly been made in this field, although it is one in which many unwarranted conclusions have been drawn, often to be refuted later by results of other workers whose experimental techniques were more careful.

As in other sections, alternating current properties, investigated by such workers as COLE and CURTIS (see ROSENE[250], 1947), ROTHSCHILD[251] (1947) and others, are omitted from the frame of reference.

BURR[47, 48] and BURR and NORTHROP[60, 61] have strongly urged the view that there exists in living organisms an electrodynamic field and that 'the recorded potential differences are measures of a store of electrical energy which can be drawn upon by biological activity and, hence, they provide valuable evidence of the nature and range of those activities. Moreover, the pattern of the potential differences defines an electrodynamic field whose forces critically influence growth and development and determine the pattern of living things'. This view has not won support from the vast majority of those working in the field. The simultaneity of certain physiological and electrical changes discussed below suggests a connection between the two, but does not by itself imply that either one is caused by the other.

Growth and development— EVERS[107] (1947) has given a good account of the history of the investigation of the effect of electric currents on plant growth, from BERTHOLON'S[18] (1783) 'electro-vegetometer' onwards. This instrument supposedly collected atmospheric electricity by means of an antenna arrangement, and passed it through plants growing in the field. BRIGGS[37] *et alii* (1926) have also written a critical review of the subject and described their own experiments, which were carried out with more careful controls than most of their predecessors, and showed no increase in growth caused by electricity. (See EVERS[107] (1947) for discussion of recent positive results, also pp 128-130.)

BURR and his colleagues have shown that the potential differences exhibited by developing eggs and embryos are so oriented that the direction of greatest potential gradient is related to their morphology

(BURR[46], 1941, BURR and HOVLAND [54, 55]). Changes in potential difference have been correlated with cell differentiation in *Obelia* (BURR and HAMMETT[53], 1939), in corn roots (BURR[46A], 1942) and in *Amblystoma* (BURR and BULLOCK[50], 1941).

The effect of electric currents on the potential gradient along onion (*Allium cepa*) roots was investigated by HOYT[151] (1947). LUND, MAHAN and HANSZEN[193, 194] and EVERS and LUND[108] showed that electric current along onion roots could inhibit growth. *Figure 1* gives some of their results[108]. From 1½ to 3 hours after the commencement of measurements an electric current was passed through the apical 1·5 mm segment, producing a potential gradient (electric field) along the root. Each curve in *Figure 1* represents the average of about four experiments. Currents giving a potential gradient of 65 to 70 mv/mm in either direction did not apparently affect the rate of growth, but with a potential gradient of 75 mv/mm in the opposite direction to the natural pd (•—•) the rate of growth was decreased reversibly. With 100 mv/mm recovery was much less marked. Potential gradients of 75 and 100 mv/mm which enhanced the natural pd (o---o) had little effect on the rate of growth.

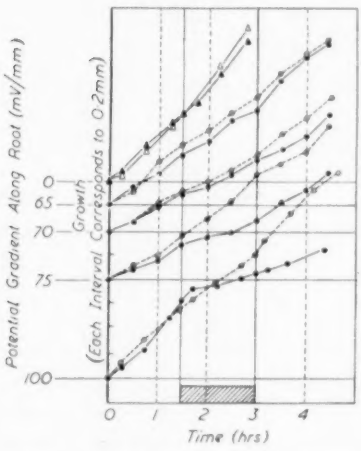


Figure 1. Effect of electric current on the rate of growth of the apical 1·5 mm segment of onion root

In other experiments LUND, MAHAN and HANSZEN[194] obtained reversible inhibition of growth with about 37 mv/mm applied to the whole root and irreversible inhibition with about 42 mv/mm. These inhibitory potential gradients were in the opposite direction to that produced by the root itself, and at least ten times as great. The inhibitory effect of electric current on growth was not confined to one segment of the root ; both the region of active cell division and that of cell elongation were effected.

There have been recent reports of the acceleration of growth by electric currents (EVERS[107]) as well as of changes in curvature (WENT and THIMANN[277] (1937) ; also pp 128-130). CHOLODNY and SANKEWITSCH[71] (1937) found that the rate of growth of *Avena* coleoptiles was increased by about 30 per cent when an electric current was passed through them from base to apex.

These results obtained with plants are of interest in relation to the

effects of applied electric currents on the rate of acid secretion by gastric mucosa discussed later.

Anaesthesia— It is not proposed to deal at length with the effect of anaesthesia or narcosis on the potential differences developed by living tissue. They have recently been studied by BISHOP[24] (1932) and by HARMAN[136] (1941) who gave references to work by Burge and Burr and their respective co-workers. BURGE's electric theory of anaesthesia[44, 45] has not been generally accepted, and his experimental work has also been criticized. It seems clear that some types of anaesthesia produce changes in the pd's developed in the intact animal, and much evidence is available as to the effects of various narcotics on tissue permeability (HÖBER[142]) which is intimately related to the potential differences developed.

Ovulation—In 1935 Burr[52] *et alii* published a note on the detection of ovulation in the intact rabbit. The animals were mated, and the pd subsequently recorded between the vagina and the abdominal surface. Several surges of 7 to 20 mv were observed about 10 hours after mating (when ovulation usually occurs). Subsequent laparotomy showed the same number of ruptured follicles in each rabbit as the observed number of potential surges.

During the next ten years much experimental work was done and a number of papers published on the timing of ovulation by pd determinations. The experiments fall into two main classes : *a* the timing of ovulation to within a few minutes by measurements from some part of the genital tract and *b* the timing of ovulation to within about a day by measurements from external surfaces of the body. There is a certain amount of overlapping between these classes.

Class *a* includes confirmation by REBOUL, FRIEDGOOD and DAVIS[235] (1937) of the results quoted above for the rabbit. BURR, MUSSELMAN, BARTON and KELLY[58, 59] published rather similar results (a surge of about 70 mv in the pd between the vagina and skin above the symphysis pubis) for a woman, the occurrence of ovulation being confirmed by subsequent laparotomy.

Measurements throughout the whole oestrous cycle of rats were carried out by ROGERS[247–249] and by NICHOLAS and CARMOSINO[221] (1944). The evidence suggests that in rats the vagina becomes more negative with respect to a point on the body surface (by perhaps 20 mv) during oestrus than in dioestrus, and that, some hours after the first indication of oestrus, a number of sharp pulses (5 to 40 mv) occurs in the pd, whose direction is not constant in different animals. If the ovaries and tubes are examined immediately after the pulses cease, recent eggs are recovered, their number bearing no relation to the number of pulses.

All these experiments are open to the criticism that the potential differences recorded may have been affected by *p*H changes of the vaginal or cervical fluid. These would affect the background pd level rather than the rapid pulses observed. Simultaneous electrical and plethysmographic records have shown that electrical fluctuations are associated with muscular activity in the uterus and vagina (BOLING and BURR[33], 1941), and it seems likely that the pd pulses are the result of uterine contractions which accompany ovulation and which, once initiated, continue and are transmitted to neighbouring parts of the uterine and vaginal walls. BOURDILLON[34] (1939) made a thorough investigation of the pd between the vagina and abdominal surface in rabbits, using a galvanometer with 1/30 sec period, and found that contractions of the vagina (induced for instance with ergometrine) were accompanied by pd pulses, the active portion being negative while contracted. Pd measurements have been made directly on the ovary by REBOUL, DAVIS and FRIEDGOOD[234] (1937) who used rabbits, and by SNODGRASS, ROCK and MENKIN[255] (1943) whose experimental material consisted of three women in the course of laparotomy ' within the patient's estimated ovulation time '. The former group of workers was able to reproduce the potential change at ovulation measured from the vagina by puncturing the ovary ; on the other hand the second group observed no greater pd than 2 mv.

Before considering the evidence under class *b*, the approximate timing of ovulation by measurements from external surfaces of the body, some intermediate experiments should be considered. LANGMAN and BURR[186] found that in humans the cervix became more negative during one or two days at the estimated time of ovulation, and also sometimes at the end of menstruation. SNODGRASS, ROCK and MENKIN[255] suggested that this negative shift was due to an increase in the *p*H of the cervical fluid. They also found some relation between the pd and the temperature difference between the electrodes in the vagina and above the symphysis pubis.

The reports that ovulation could be timed to within a day by simple measurements from two accessible parts of the body attracted a considerable amount of attention, owing to the practical importance of a rapid and reliable method for humans. ALTMANN[1] (1941) published positive results for the sow, and FINCH, YERKES and ELDER[113] (1937) for the chimpanzee. The successful timing of ovulation in women by daily measurements of the pd between fingers of the right and left hands was reported by BURR and MUSSELMANN[56, 57] and confirmed by PARMENTER[227] (1939) and others. Positive, but more cautious, results were published by BARTON[15] (1940). The published evidence was roughly that, when disturbing factors such as slight abrasions had been

eliminated, the steady pd between the index fingers of the right and left hand was several times greater during one or two days near the middle of the menstrual cycle than on any other day, and in some women the pd curve was very similar for a number of consecutive cycles. This cyclic variation was absent in men and in non-menstruating women. In some experiments the date of ovulation could be identified by other methods, and in general it agreed with the date estimated from the pd graph.

In 1943, however, SNODGRASS, ROCK and MENKIN[255] published a paper on the validity of 'ovulation potentials', in which they described various careful sets of experiments designed to determine factors affecting the pd between two fingers, and their results gave no evidence of cyclic potential changes between finger-tips that might be related to the occurrence of ovulation. BURR[49] (1945) published a reply to this paper in which he maintained his ground. CRANE[83] (1944) found that while pooled results from daily measurements over about one hundred menstrual cycles might be related to the occurrence of ovulation, individual results could not be relied upon (*cf* remarks on carcinogenesis on p 110). The distribution throughout the menstrual cycle of an abnormally high pd between homologous fingers had some similarities to distributions of the time of ovulation estimated by various methods (*e.g.* KURZROK[185], 1937). However, individuals gave widely different results; a few showed cyclic variations, but most showed none when studied individually.

The general conclusion seems to be that while in some cases consistent results are obtained, this is by no means always so, and the method is not an unequivocal one for the clinical determination of the date of ovulation, as was at one time hoped. Very little is known about the variations in the conditions and behaviour of the skin in relation to the menstrual cycle. It has been reported by KLAUS[172] (1927) that the choline content of sweat is much greater at or near menstruation than in the intermenstruum. ASHLEY-MONTAGUE[7] (1940) has given a review of other variations known or suspected to occur. A proper evaluation of the results discussed above must await developments in physiology and biochemistry.

Healing of wounds— Just over a hundred years ago DU BOIS-REYMOND[103] demonstrated the existence of a pd between injured and uninjured parts of nerve and muscle, the injured portion being negative in an external circuit with respect to the uninjured portion. Since then a great deal of work has been carried out on injury potentials in nerve and muscle, plants, frog skin and other tissues. The results have been reviewed by BEUTNER[22] (1944), HEILBRUNN[138] (1947), HÖBER[142] (1946) and others.

It is proposed here to deal only with electrical changes accompanying the healing of wounds. *Table VII* gives representative results for wounds in which the skin has been pierced. In all cases cited the wound itself became positive with respect to the neighbouring intact

Table VII. Potential Differences Associated with Skin Wounds During Healing

Species	*Tissue*	*Wound*	*Points between which pd measured*	*Sign and variation of pd*	*Average range of variation in pd (mv)*
Guinea-pig	*right flank*	*vertical incision through skin and subcutaneous fascia (sutured)*	*between site of wound and neighbouring point on skin*[81]	*positive for about 4 days, then negative until healing complete (say 16th day)*	+ 9, − 18
Wistar rat	*dorsal skin*	*superficial cut*	*between site of wound and neighbouring point on skin*[86]	*positive on same and next day*	+ 11
Human	*skin surface*	*for herniorrhaphy, celiotomy and laminectomy*	*between point in immediate vicinity of wound and point at some distance from it*[84]	*positive for about 4 days, then negative, re-approaching zero at about 14 days*	+ 3, − 2
Human	*thumb*	*cut 4 mm deep (by broken glass)*	*between wound and ball of thumb*[82]	*positive until wound healed (8 days) then negative*	+ 40, − 20
Human	*finger tips*	*sand paper abrasion*	*between injured finger and intact homologous finger*[10]	*positive 'until wound closes over'*	+ 15 to + 50

skin. In general the value of the pd depended rather little on the type of wound, or its depth or extension, so long as contact was established with the inner surface of the skin. Those measurements which were continued for long enough show that as the wound healed over, the site of the wound became less positive, finally reversing its sign when the healing was complete.

Figure 2 shows a fairly typical pd-time curve (CRANE[83], 1944). A clean wound about 4 mm deep was made accidentally near the base of the thumb, and the pd plotted in *Figure 2* is that between the wound and the ball of the thumb, measured with 0·16M NaCl-KCl bridge—Ag/AgCl electrode systems. The pd rose from 30 mv to 40 mv during the first 24 hours, and after remaining at about this level for four days, dropped rapidly as the wound healed over, becoming negative nine days after wounding when healing was complete. Thereafter the change was much slower, but it continued for another ten days.

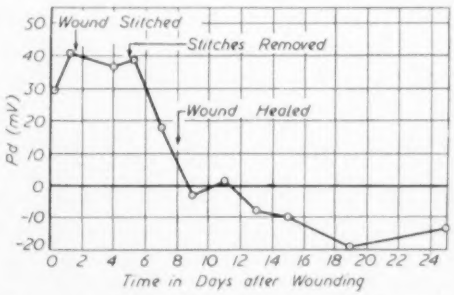


Figure 2. Variation of mean pd between a deep wound at the base of the thumb and neighbouring uninjured skin during healing

The measurement of potential changes has been used to estimate the acceleration in the rate of healing under different treatments (BARNES[10], and others).

REHM[237] studied the potential changes following injury to the mucosal surface of the dog's stomach. The injuries were made with ethyl alcohol, ether or boiling Ringer solution. In every experiment the injured portion became 10 to 90 mv more positive with respect to a neighbouring uninjured portion of the mucosal surface, which is negative with respect to the serosa.

Carcinogenesis— The early diagnosis of cancer provides an attractive field for almost any newly developed biological method, and it is not surprising that attempts have been made to discover a relation between carcinogenesis and potential differences developed. Unfortunately early suggestions of success (BURR, SMITH and STRONG[62], 1938) have not been substantiated (BURROWS, ROE and SCHOBER[66], 1945).

The first group of workers, referring to Strong strain A mice having spontaneous adenocarcinoma of the mammary gland, say that 'in those mice in which cancer appeared before the 260th day, marked rises amounting to several thousands of microvolts were encountered in the readings across the chest. . . . The voltage rise across the chest appeared in some instances from ten days to two weeks before the tumor was detected by palpation'.

The second group of workers induced tumours in Medical Research Council mice with the carcinogen 9,10-dimethyl-1,2-benzanthracene. They give a statistical analysis of their results, which shows the existence of certain significant differences in the pd and its variance. They conclude, however, that 'as almost all the significant changes were observed in the pooled experimental measurements and not in individual animals, this technique cannot be used at present for the early diagnosis of cancer'. BURR, STRONG and SMITH[63] also conclude in one paper that there is not a specific diagnostic response. At present, therefore, electrical measurements do not offer a satisfactory diagnostic method for carcinogenesis; whether they may do so when more is known and understood about the pathological changes themselves, remains to be seen.

SPECIFIC ELECTRIC ORGANS

Distribution— It was suggested earlier that the torpedo played an important part in the development of fundamental physics in the eighteenth century, and that the morphology of its electric organ was partly responsible for the design which Volta adopted for his pile at

the close of that century. During the nineteenth century the various species of electric fish played a less significant part in the development of either physics or electro-physiology, and it was not until the twentieth century was well advanced that they again became important.

During the past ten years COATES, COX, NACHMANSON and others have brought to light important information concerning the exergonic and (electrical) endergonic processes in electric organs, and the relations between them. An endergonic process is defined as one in which the free energy of the reactants increases ; energy is required to maintain such a process. An exergonic process is one in which the free energy of the reactants decreases ; energy is liberated and may be available to drive an endergonic process. HINSHELWOOD[140] (1947) discusses possible mechanisms of coupling. In living tissues the overall reactions are exergonic, although some of the steps are endergonic (CORYELL[76], 1940).

So far as is known the defensive and offensive use of electricity in the animal kingdom, for immobilizing enemies and stunning or killing prey, is confined to a few fishes and elasmobranchs, although there have been some unsubstantiated records for insects and for a snail. In these animals the problem of converting chemical energy into electrical energy available externally, in some cases at the rate of several kilowatts, was solved many millions of years before man devised inorganic systems to do so. The most important species are described briefly below ; further details are given in *Table II.*

The electric eel (*Electrophorus electricus, formerly called Gymnotus electricus*)—This is the most highly specialized and most formidable of all electric fish, developing up to 600 volts on open circuit. It is almost defenceless apart from its electric power, and even its eyes are poorly developed. The electric eel may reach more than 2·5 m in length, but it is not as heavy as the flat *Torpedo occidentalis*. It inhabits the water and marshes of the Orinoco and Amazon, but has proved very useful for experimental study in other parts of the world because, in contrast to other electric fish, it lives well in captivity and can even be kept out of water for some minutes without any reduction in its electrical activity. Moreover its electric organs are easily accessible. There are three pairs of these which together comprise half the weight of the fish : the main organs which extend along the posterior four-fifths of the fish and are responsible for the high voltage output quoted, the much smaller organs of Hunter which are under the entire length of the main organs and seem to discharge with them, and the organs or ' bundles ' of Sachs lying above the main organs in the posterior half of the fish. These discharge separately from the main organs and

develop only about 40v (Cox and Coates[79], 1938 ; Coates and Cox[72], 1939 ; Cox[77], 1943 ; Chagas[69], 1947).

Torpedo— About twenty species of the genus *Torpedo* have electric organs developed to some extent. In each there is a single pair of organs, which extend from the dorsal to the ventral skin, lying in the flat body of the fish just outside the line of gill slits. They usually comprise about one sixth of the total weight of the fish (Fritsch[121], 1890).

Torpedo occidentalis— This is the largest of all electric fish, reported to reach over 100 kg. It is found on the North Atlantic coast of the United States. The fish can produce an open circuit pd of 200 v or more, but the pd is very much less than this if it is in poor condition. It is likely that a pair of organs from a single fish can produce a peak current of 50 amps, and a maximum power of 6 kw (about 1·5 watts per g of electric tissue).

Other species of Torpedo— *Torpedo marmorata*, found in the warm seas of southern Europe, was studied in some detail by Gotch[127, 128] (1887, 1888), and more recently by Fessard[112] (1946). Hunter's[158] (1773) anatomical study was made at the request of Walsh on torpedoes caught in Torbay in 1773. The larger one was found to weigh ' 53 lb avoirdupois (24 kg) and to measure four feet (1·2 m) in length, two feet and a half (0·8 m) in its extreme breadth and four inches and a half (11 cm) in its extreme thickness ' (Walsh[276], 1774).

The electric ray (*Narcine brasiliensis* {*Ölfers*})— This fish has less well developed electric organs. Those studied by Cox and Breder[78] (1943) had an average weight of about 270 gm, one sixth of which represented the single pair of electric organs. The peak open circuit voltage was 14 to 37 v, and the maximum power output was never more than 35 watts (about 0·5 watts per gm of electric organ). The electric organ is well developed at birth, after which growth consists in an enlargement of the electroplaxes rather than an increase in their number. Narcine shows marked morphological and electrical similarities to torpedo, and behaves electrically like a smaller edition of it (Breder and Springer[36], 1940 ; Cox and Breder[78], 1943).

Other rays— Various species of *Raja* are of special interest in connection with the development of the electric organ. In the small and primitive starry ray *Raja radiata*, the electric organ is situated in the tail, and although extremely small and apparently functionless, it is in a state of progressive development : there is no indication of degeneration or any evidence that a higher stage of development of the electric organ was ever reached in this species (Ewart[109], 1888). Other species with electric organs more fully developed include the undulate ray *Raja undulata* (Auger and Fessard[8], 1939), the thornback

ray, *Raja clavata* (KRAJUCHIN, BODROVA and MOGILA[176], 1938), the skate *Raja batis* and the sandy ray *Raja circularis* (EWART[109, 110], 1888, 1892). The electric organ, situated in the tail of most skates and rays, is in general poorly developed, and it is not often realized that such an organ exists in so many British species.

The electric catfish (*Malapterurus electricus*)— This fish, which still lives in the rivers of tropical Africa and the Nile, is pictured on the walls of early Egyptian tombs, together with fishermen who must have known its electric power. It grows to three feet (1 m) in length, although the specimens examined by GOTCH and BURCH[129] (1896) were only 5 to 6 inches (13 to 15 cm) long. These were caught at the mouth of the river Senegal, and were kept in tanks ' in excellent health for a period of over six months '. The emf developed by 10 cm of excited organ was estimated, from the strength and character of the shock it gave, to be at least 75 v, corresponding to 40 mv per electroplax (FRITSCH[121], BETHE[19]).

The electric star gazer (*Astroscopus*)— In three species of this genus portions of one or more of the eye muscles have been converted into an electric organ, occupying the enlarged orbit in which the small eye has been crowded forward and stares permanently upwards (WALLS[273], 1942). This fish buries itself up to the eye in sand and produces an electric shock which results in some other small animal falling into its cavernous mouth. A number of Mormyridae and a few other fishes also have electric organs.

Structure of electric organs— In the electric catfish, *Malapterurus electricus*, the electric organ is derived from the integument, and consists of a greasy or gelatinous mass surrounding the body from head to tail (HEILBRUNN[138], 1947). In other species the organs are metamorphosed muscular fibres with their nerve endings and motor end plates.

The electric tissue in the electric rays and electric eel is similar, and is different from any other tissue. In the torpedo it consists of a large number of small columns or prisms, each extending from the positive (dorsal) surface of the organ to the negative (ventral) surface. Each column is composed of a pile of electric plates or electroplaxes. In the eel the electric organs are divided by horizontal septa into layers which correspond roughly to the columns of the organ in the rays. These layers are divided vertically by thin septa into a pile of electroplaxes. In all species there is a branched nerve ending on the negative surface of each plate.

The electroplax is thus the unit of the ' battery ', and in all species studied the pd per electroplax has been about 50 to 150 mv *i.e.* of the same magnitude as other bioelectric potentials. The electric organ

functions as such because it contains a large number of units in series and in parallel. Du Bois-Reymond[103] (1848) first pointed out that the array of electroplaxes is relatively more in series in the eel and more in parallel in the rays, as it would be if the resistance of the electric organ were made to match the resistance of an external circuit through the water, since the eel lives in fresh water and the rays are marine species. Thus *Torpedo occidentalis* may produce a current of 50 amps, but the current from the electric eel is not more than about 1 amp (*cf Figure 3*, in which resistances are given for two rays and part only of the electric organ of the electric eel). The long body of the eel and the flat body of the rays are both suited for developing electric power in their own media.

In the rays the electroplaxes are of about the same thickness all along the column, but in the electric eel this is not so. As the cross section of the organ decreases towards the tail of the fish, the thickness of each electroplax layer increases, in such a manner that the volume per electroplax layer is approximately constant. Although the thickness of the electroplax layer varies by a factor of ten, the maximum current density and the resistance per unit area at peak discharge are also constant for each electroplax layer at different distances along the length of the fish (Cox, Coates and Brown[82], 1945).

For all species studied the peak voltage during a discharge through various external resistances is linearly related to the peak current in them. *Figure 3* shows pd current graphs for electric organs of three species. The linearity of the curves indicates that Ohm's law applies to the organ itself *i.e.* that its resistance is independent of the value of the current. This result should be compared with that for such secretory tissues as gastric mucosa and frog skin (p 127). The resistance of the electric organ of the electric eel is constant up to a current density more than 10^3 times as great as that for gastric mucosa : it is a highly specialized organ.

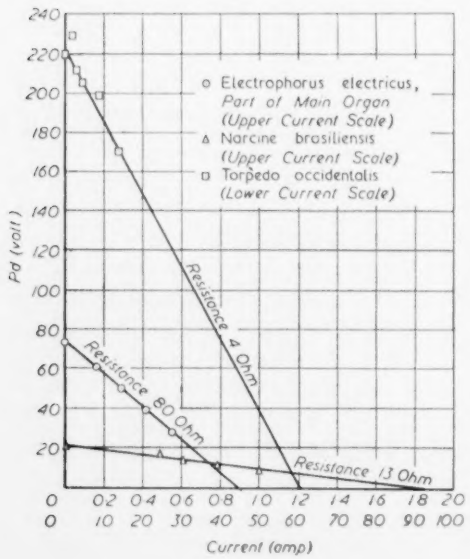


Figure 3. External peak voltage plotted against external peak current for three species of electric fish

Discharge of electric organs— The electric organs of fishes do not produce a steady power output, but operate by means of trains of discharges. Both the electric eel and *Torpedo occidentalis* throw off fairly regular trains, each consisting of a number of discharges of several milliseconds' duration, separated by intervals of a milli-second or so. COATES and COX[73] report three or four discharges per train in the electric eel and at least a dozen in *Torpedo occidentalis*. The peak voltage and current referred to earlier in this section are the maximum values during any one discharge. The peak power produced is much less than their product because they are not produced simultaneously. The maximum power estimated for both organs of one specimen of *Torpedo occidentalis* was a little over 6 kw ; the average power would be perhaps 2 kw for one discharge and half to 1 kw for a train of discharges. The relations between the energy expended electrically and that available from metabolism is discussed later.

One important fact which has recently been brought to light by COX, COATES and BROWN[82] is that the discharge of the electric organs of the electric eel is probably brought about not by the sudden development of a large emf, but by a temporary lowering of their resistance which enables a fairly constant emf to pass a large current for a short time. This interpretation has however been challenged by FESSARD[112] (1946). If the emf developed is E, and the internal resistance of the electric organ and the leak resistance in parallel with it r and R' respectively, then the external pd V, measured when the external current is I, is given by $V(1 + r/R') = E - Ir$.

Figure 4 shows V plotted against I for three instants during the discharge of the electric organ of an electric eel (after the initial rise in voltage). Since the slopes of these lines are different, either r or R' must vary during the discharge, and since their intercepts are also different either E or r must vary. Now the three lines (which are typical of many other sets) are concurrent, as they would be, at the point $(I = -E/R', V = E)$, if E and R' were constant and r were the variable. The authors therefore suggest that the variation in the

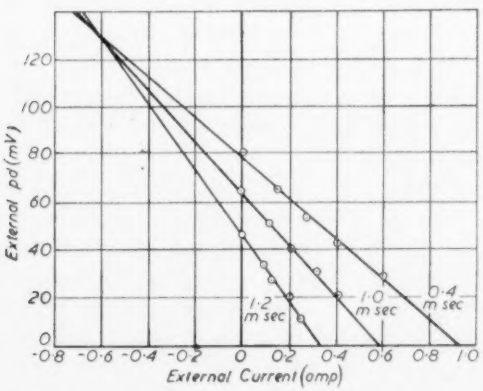


Figure 4. Relation between external pd and external current at 0·4, 1·0 and 1·2 msec after commencement of a discharge in the electric organ of Electrophorus electricus[82]

internal resistance is the primary electrical change by which the discharge takes place, the emf remaining practically constant. Except during the discharge the resistance of the electroplax boundaries is so high that no appreciable current flows. The discharge would be started by a very large and rapid increase in the permeability of the electroplax boundary to the ions, whose movement constitutes the current in the electric organ, its resistance therefore dropping rapidly. DAVIES and OGSTON[97] (1950) have given evidence that the initiation of HCl secretion in gastric mucosa apparently depends upon a sudden rapid increase in the permeability of the oxyntic cell wall to H^+ and Cl^- ions. In this case the resistance of the mucosa as a whole actually increases (CRANE, DAVIES and LONGMUIR[89], 1948). Nothing is known about the resistance offered by the various parts of the electric organ to current passed from an external source of power.

In the discharge of the electric organ of the electric eel, a pulse of potential gradient is propagated along the organ at a much higher speed than that at which nervous impulses are normally transmitted (COATES, COX, ROSENBLITH and BROWN[74], 1940). For details of this and other properties of the discharges the reader is referred to the original papers by workers mentioned in this section.

RELATION OF BIOELECTRIC POTENTIALS TO SECRETORY FUNCTION

In living tissue, cells do useful work (endergonic processes), expending energy derived from exergonic processes during the oxidation of foodstuffs. It is known that some of this energy can be stored as high energy phosphate bonds and is available as such (LIPMANN[187], 1941 ; FRUTON *et alii*[122], 1944 ; OGSTON and SMITHIES[223], 1948). For certain secretory membranes it is possible to measure both the respiratory and the secretory rates and also the potential differences developed, and thus to attempt to discover the part, if any, played by electrical potential differences in both exergonic and endergonic processes, and in the intermediate storage of energy.

This section is concerned only with active transport (KROGH[181, 182], 1937, 1946) which is an endergonic process, and not with passive permeability as in Donnan membranes. Donnan distributions have been invoked to account for continuous active secretion, but Donnan himself recently denied this possibility, condemning it as a logical contradiction (personal communication by DONNAN to HOLLANDER[149], 1943).

The high concentration gradients against which active secretion can take place make it one of the most impressive phenomena in the physical chemistry of biological tissues. For instance gastric hydro-

chloric acid is secreted by oxyntic cells as a solution whose H^+ ion concentration is several million times as great as that in the cell, and this concentration difference is maintained across a distance not greater than 1μ, and probably much less (CRANE, DAVIES and LONGMUIR[89]).

Studies of active secretion and absorption— Active secretion and/or absorption occurs in various tissues of many plants and animals (HÖBER[142]). KROGH[182], LUNDEGÅRDH[196–199] and ROBERTSON and WILKINS[245, 246] have recently discussed active transport of ions by various plant cells. KEYS[167, 168] and KROGH[182] studied active transport of Cl^- ions in the gills of fish. Absorption and/or secretion takes place in all parts of the gastro-intestinal tract, and both biochemical and bioelectrical investigations have been carried out especially on gastric mucosa (REHM[238], REHM and HOKIN[242], RICE and ROSS[243]; TEORELL and WERSÄLL[260]; CRANE, DAVIES and LONGMUIR[88, 89]; CRANE and DAVIES[87]). The acid-secreting oxyntic cells in mammalian gastric mucosa are among the most active mammalian cells known (CRANE and DAVIES[85]).

Although some insects provide suitable material for the study of secretory problems (WIGGLESWORTH[278]), they represent a relatively unexplored field, and virtually no bioelectric measurements have been made on them.

The electrical properties of nerve fibres, and active transport of ions in them during both resting and active states, have been studied recently by HODGKIN[143], HODGKIN and HUXLEY[146, 147], HODGKIN and KATZ[148] and KEYNES[165, 166].

STIEHLER and FLEXNER[258] and FLEXNER and STIEHLER[114] used dyestuffs to demonstrate maintained oxidation-reduction potentials in pig choroid plexus. By the use of acid and basic dyes, they related them to the active transport of Cl^- ions into the cerebrospinal fluid. FRIEDENWALD and STIEHLER[120] used similar methods to study the secretion of the intraocular fluid by the ciliary body; there has, however, recently been controversy as to whether a secretory process is involved (DAVSON, DUKE-ELDER and MAURICE[98], KINSEY[170]).

Skin, particularly frog skin, has been a favourite experimental material. HUF[152–157] has carried out a series of investigations on the effect of anoxia and respiratory inhibitors on the pd and secretion of frog skin, using both open sheets and tied bags of skin. He produced a wealth of experimental evidence to show that the transport of sodium chloride and of water is an active endergonic process, and that the pd across the skin is related to this transport.

Use of radioactive isotopes— By the use of radioactive isotopes it is possible to trace the movements of certain ions among others of the

same species, and this technique has obvious applications in the study of secretion and active transfer. In a recent review of radioactive isotopes as indicators in biology, SACKS[252] has collected information from nearly three hundred papers, and it is proposed here to refer only to work of special interest in connection with problems discussed elsewhere in this review. The work of KEYNES[165, 166] on nerve has already been mentioned.

HOLM JENSEN[150] used Na^{24}, K^{42} and Cl^{38} to determine the Na^+, K^+ and Cl^- uptake in *Daphnia magna*. JØRGENSEN, LEVI and USSING[160] used Na^{24} to determine the Na^+ uptake in the axolotl, and USSING[265, 266] used Na^{24} and Cl^{38} to investigate the influx and outflux of Na^+ and Cl^- ions through isolated frog skin, and measured the accompanying pd across it. VISSCHER *et alii*[270] (1944) used Na^{24}, Cl^{38} and also D^2 in an investigation on the simultaneous movement of water and Cl^- and Na^+ ions through the intestine, between the ileal lumen and the blood, in anaesthetized dogs. The observed transfer rates were far higher than those expected from simple diffusion. COPE, COHN and BRENIZER[75] found that Na^{24+} was absorbed about one hundred times as rapidly (per unit area) from the pyloric antrum as from the body of the stomach, and that this high absorption rate continued during HCl secretion.

GREENBERG, *et alii*[132] showed, by the use of Na^{24}, K^{42}, Rb^{86} and Sr^{89}, that the capillaries of the choroid plexus are as freely permeable to these cations as those of the tissues generally. KINSEY *et alii*[171] used various isotopes in experiments on the production of aqueous humour, the results of which indicate that osmotic work is involved (see also p 117).

One importance of tracer studies is that by comparing the variation in the measured flux through a membrane (when the concentrations of the surrounding solutions are varied) with that calculated for diffusion, information may be obtained as to whether or not the transport of a species of ion is endergonic. One application, as yet little developed, is the determination not only of the relative number of ions passing through a membrane under different conditions, but also of the absolute number. This is of value in building up a picture of the events happening within the membrane in terms of the individual ions which take part in them, and in working out the energy relations involved.

Potential difference across secretory membranes— The potential differences across secretory membranes are among the highest developed across any living membrane (see *Table II*). It seems clear, moreover, that these potential differences are related to the secretory process, and cannot be explained on a simple physical basis.

The next section shows that the pd across those secretory membranes which have been studied (frog skin, isolated frog gastric mucosa and dog stomach) is also capable of maintaining a continuous electric current through an external circuit, which involves the expenditure of a considerable fraction of the energy available from metabolism. It would be useful to know whether this is true of all secretory membranes, and whether it is true of tissues such as muscle and nerve which expend energy in other ways.

Gastric mucosa, which secretes hydrochloric acid (0·12 M in amphibia, *c.* 0·16 M in mammals), offers considerable advantages for the simultaneous study of electrical and secretory activity, and has recently been used fairly widely, for amphibia by MOND[210], DELRUE[100], TEORELL and WERSÄLL[260], and CRANE, DAVIES and LONGMUIR[88, 89]; for rats by PATTERSON and STETTEN[228]; for cats by MISLOWITZER and SILVER[208], MISLOWITZER, SILVER and ROTHSCHILD[209] and SARRE[253]; for dogs by TITAEV[262], VENCHIKOV[268, 269], REHM[240], REHM and HOKIN[242] and RICE and ROSS[243]; and for humans by SWYNGEDAUW[259], ADAIR and GOODMAN[3], QUIGLEY, BARCROFT, ADAIR and GOODMAN[233], TITAEV[262], GOODMAN[126], REHM and ENELOW[241] and RICE and ROSS[243].

Not all the work referred to above is of equal significance or even validity; with intact animals (including humans) it is extremely difficult to control experimental conditions and to interpret results. The recent experiments on dogs were done with refined techniques and their results are more reliable. With amphibian tissue, which can secrete hydrochloric acid *in vitro*, it is easier to control experimental conditions.

Figure 5 gives representative results for the change in the pd between the mucosa and serosa at the onset of acid secretion induced by histamine, and when acid secretion is stopped by the presence of thiocyanate. *Figure 5*a refers to the stomach of a dog anaesthetized with sodium amytal[240]. *Figure 5*b refers to isolated frog gastric mucosa[89]. The scales are different, and in *Figure 5*a the dry weight of the tissue was unknown so the Q_{HCl} (µl per mg dry wt per hr) could not be calculated. The ordinate is however proportional to Q_{HCl}. Both pd graphs give the pd between the secretory and nutrient sides, which decreased fairly sharply when histamine-induced acid secretion started and rose again (although not quite to its previous level) when this was inhibited by sodium thiocyanate. The rise in pd during the first half hour after mounting (*Figure 5*b) was usual[89].

It seems certain, from experiments with different ions in solution and with models, that the pd change on secretion is not a simple diffusion pd due to the change in H^+ ion concentration; wide variation in the ionic composition of the secretory solution had little effect on the pd (REHM[240], TEORELL and WERSÄLL[260], REHM and HOKIN[242], RICE and ROSS[243]; CRANE, DAVIES and LONGMUIR[89]).

The HCl secreted by gastric mucosa is produced in the pericanalicular zone of the oxyntic cells. On the basis of the evidence available, Crane, Davies and Longmuir have suggested an electron-cycle mechanism for the utilization of energy available from metabolism,

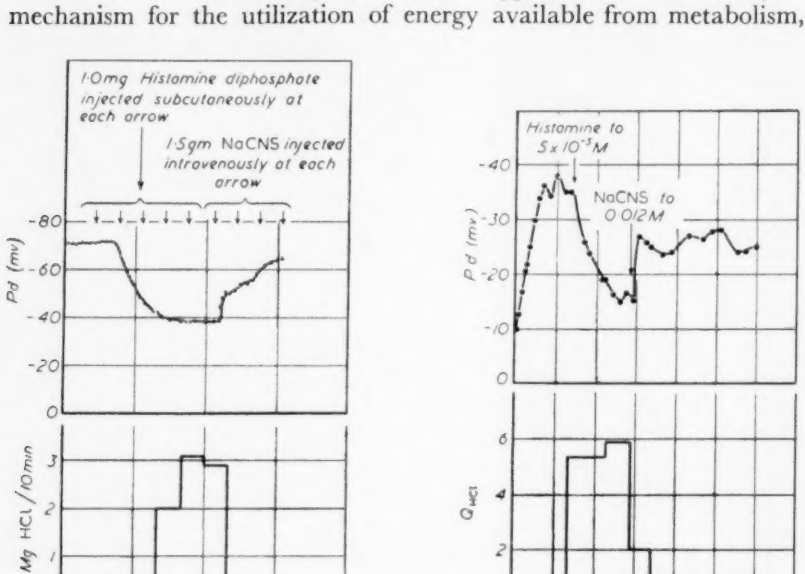


Figure 5. Variation of pd across **a** *dog stomach and* **b** *isolated frog gastric mucosa, and of the rate of* HCl *secretion*

possibly as high energy phosphate bonds (see BALDWIN[9]), in an electrochemical process which results in a net separation of the electrically charged H^+ and OH^- ions. A possible supplementary process for the production of H^+ and OH^- ions by oxyntic cells (suggested independently by ROBERTSON[244]) is the reaction of O_2 with substrate H atoms, *via* cytochrome oxidase and the cytochromes, to form H^+ and OH^- instead of H_2O (CRANE and DAVIES[85]). Evidence as to the energies involved are discussed in the next section.

Pd-current characteristics of secretory membranes— One electrical property which seems to be common to secretory membranes, but absent in those studied which have no secretory function, is the S-shaped pd-current characteristic, the slope of which gives the resistance as a function of current density in the tissue. A typical characteristic for isolated frog gastric mucosa has been published elsewhere (CRANE, DAVIES and LONGMUIR[90]). *Figure 6* gives characteristics, not previously

published, obtained by CRANE, DAVIES and KHALIL in 1948 during an experiment on the isolated swim bladder of the perch, *Perca fluviatilis*. In the perch there is no connection between the lung and the swim bladder, which contains in the lining epithelium patches of glands capable of secreting oxygen and carbon dioxide. KHALIL[169] (1937) has given a general discussion of swim bladders ; see also BLACK[25, 26].

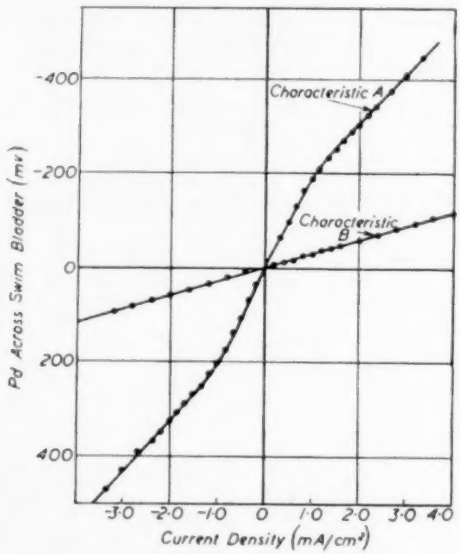


Figure 6. Pd-current characteristic of perch swim bladder : A part containing gas gland B part not containing gas gland

Characteristic A of *Figure 6* shows the variation with current density of the pd across a portion of the swim bladder containing the gas gland; characteristic B was obtained with a portion of the same swim bladder not containing the gas gland. The resistance of 1 sq cm of perch swim bladder is given by the slope of the characteristics. For the part containing the gas gland (characteristic A) the resistance was constant at low current densities and again constant at high current densities, but at about one half its value for low current densities. For live isolated frog gastric mucosa the ratio is about one third.

Characteristic A is similar in shape to those obtained for other live secretory membranes studied, which are listed in *Table VIII*. It should also be compared with curves for frog skin obtained by PUMPHREY[231, 232]. Characteristic B is linear, like those for other non-secretory membranes, and for dead secretory membranes ; the resistance was constant for all current densities used.

The swim bladder of the perch was chosen for illustration because it was possible to obtain characteristics for secretory and non-secretory portions of the same membrane, and because it shows that the S-shaped characteristic is not confined to membranes secreting a fluid, but is also obtained from one which secretes gas.

Table VIII shows results for isolated secretory and non-secretory membranes (CRANE and DAVIES[87], 1950). The membranes were bathed with a nutrient solution and/or isotonic NaCl solution ; the nutrient solution was that of KREBS and HENSELEIT[178] (1932), if

necessary made isotonic by dilution, and containing 0·2 per cent glucose. The pd across the membrane was measured by means of saturated KCl bridges and calomel electrodes.

The secretory side of all amphibian secretory membranes studied was negative in an external circuit with respect to the nutrient side. All such membranes mounted alone (without their muscular coat) had S-shaped pd-current characteristics, with three linear portions. The three duodenal and two rectal mucosae were mounted with the non-secretory muscular coat attached, and of these all except one duodenal mucosa gave an S-shaped characteristic.

Of the three isolated mammalian gastric mucosa listed in *Table VIII* polecat mucosa was the only one which secreted acid, and the only

Table VIII. Electrical Characteristics of Secretory and Non-secretory Membranes[87]

Membrane (all isolated)	*Number studied*	*Number giving S-shaped curve*	*Number giving linear curve*	*Temperature °C*	*Approximate pd across membrane (mv)*
Secretory:					
Amphibian					
gastric mucosa	39	39	0	25	−30
pyloric mucosa	2	2	0	25	− 5
duodenal mucosa (+muscle)	2	1	1	25	− 2
oesophageal mucosa ..	1	1	0	25	− 2
rectal mucosa (+muscle)	1	1	0	25	−17
skin	5	5	0	25	−34
Mammalian					
polecat gastric mucosa ..	4	4	0	25, 38	−10, −21
hedgehog gastric mucosa ..	2	0	2	25, 38	0, 0
rat gastric mucosa ..	1	0	1	25	0
Fish					
perch (Perca fluviatilis) *swim bladder (with gas gland)*	2	2	0	25	0
Non-secretory:					
Amphibian					
dead gastric mucosa ..	28	0	28	25	0
stomach muscle layer ..	3	3*	0	25	0
Mammalian					
rat diaphragm	7	0	7	38	0
rat kidney capsule ..	1	0	1	25	0
guinea-pig atrium ..	2	0	2	38	0
Fish					
perch (Perca fluviatilis) *swim bladder (no gas gland)*	1	0	1	25	0
Cellophane	1	0	1	25	0

* Slight breakdown at 5 ma/sq cm

one which developed a pd. All four specimens had S-shaped pd-current characteristics. Isolated rat and hedgehog gastric mucosae did not secrete acid, developed no pd, and gave a linear characteristic through the origin.

No non-secretory membrane studied exhibited a pd, and all gave linear pd-current characteristics (*Table VIII*) even up to current densities many times the breakdown current density of secretory membranes.

Thus every secretory membrane which was known to secrete under the conditions of the experiment, except one duodenal mucosa which developed only a low pd, had an S-shaped characteristic, and every non-secretory membrane had a linear characteristic over the same range of current densities. It seems that a constant resistance at low current densities and a considerably lower (constant) resistance at higher current densities, together with an intermediate ' breakdown ' at a fairly well defined current density, may be a specific property of tissues containing secretory cells (CRANE and DAVIES[87], 1950). Results are needed for other secretory membranes, and for those which expend energy in other ways.

It has already been shown that the pd-current characteristic of the electric organ is linear over the whole range of the current densities produced by its own discharge, in some cases up to several amps per sq cm (one direction only). No results are available for applied currents, either for the discharging or resting electric organ.

Although secretory membranes are essentially asymmetrical, both in structure and function, no well-defined rectification was found with any of the membranes listed in *Table VIII*, although some specimens showed slight rectification, which might be in either direction. Nerve fibres (squid giant axon) on the other hand, show considerable rectification, permitting currents to flow more easily outward than inward (GUTTMAN and COLE[135], 1941 ; GUTTMAN[134], 1947). These authors found that the rectification and the resting potential decreased with time after dissection at the same rate, and that they disappeared concurrently. The fact that there is a difference between secretory membranes and nervous tissue is not unexpected in view of their different structure and function, but the absence of consistent rectification in secretory tissue is rather surprising.

BIOELECTRIC POTENTIALS AS A SOURCE OF ENERGY

Relation between bioelectric potentials and energy available from metabolism—There is ample evidence that bioelectric potentials are not maintained in the absence of an adequate supply of oxygen, and that they are decreased or abolished by respiratory inhibitors. Oxygen starvation

has been found to reduce the pd across frog skin to zero, the effect being irreversible after 3 to 5 hr with *Rana pipiens* (LUND[190]) or after 20 to 40 min with *Rana temporaria* (FRANCIS and PUMPHREY[119]). It is possible that the differences are due to different material, or to the less thorough exclusion of oxygen in some cases. FRANCIS[117] (1934) found that in most skins the pd decreased only when the oxygen concentration of the bathing solutions was reduced below 20 per cent (see also HUF[155]).

CRANE, DAVIES and LONGMUIR[88, 89] found that anaerobiosis abolished the pd across isolated frog gastric mucosa, but that the pd regained its previous value when oxygen was supplied again, even after several hours of anaerobiosis. REHM and HOKIN[242] and RICE and ROSS[243] have shown that the pd across a dog's stomach wall depended upon an adequate oxygen supply to the tissue.

Similar effects have been observed with plants; for instance BLINKS, DARSIE and SKOW[30] found that the pd of *Halicystis* was reversibly inhibited by anaerobiosis.

Reduction of the pd by respiratory inhibitors has been widely observed, but the pd may also be reduced by drugs which do not affect the respiration. FRANCIS and GATTY[118] (1938) have discussed the effect of numerous respiratory inhibitors on the pd and respiration of frog skin ; they say that ' at present no single substance has been found to lower the oxygen uptake over a period of 1 hour or more and not to reduce the potential '. Similar results have been obtained for other animal structures and for those of plants (see ROSENE[250], 1947). While the processes by which energy derived from oxidative metabolism is used to maintain a steady pd in any tissue may be inhibited without affecting the initial oxygen uptake, the pd cannot be maintained in the absence of this exergonic process. On the other hand a comparison between different tissues (see *Table II*) shows that the value of the optimal Q_{O_2} of a tissue is little guide to the magnitude of an observable pd in it. Indeed tissues which have very high Q_{O_2} values, such as liver, have no membrane suitable for pd determinations.

The electric organ has proved to be useful material for the correlation of biochemical and bioelectrical activities. It has been shown that the pd developed by different portions of the electric organ of *Electrophorus electricus* and several other species runs parallel to the cholinesterase activity (NACHMANSOHN, COATES and COX[215], 1941 ; NACHMANSOHN, COX, COATES and MACHADO[217], 1942 ; NACHMANSOHN, COATES and ROTHENBERG[216], 1946). Cholinesterase is an enzyme which hydrolyzes acetylcholine ; the first recorded evidence of a possible relationship between it and the electric discharge from fish was the discovery in 1937 that the electric organ of *Torpedo marmorata* could hydrolyze 2–3

times its own weight of acetylcholine in an hour ; this has subsequently been found to be true for other species.

The resynthesis of acetylcholine requires energy, and NACHMANSOHN *et alii*[218] calculated that during 1,600 discharges of the electric organ of *Electrophorus electricus* the energy available from high-energy phosphate bonds by the breakdown of phosphocreatine was 3–4 times the external electrical energy expended. This result was based on the value 10^4 gm cal for the energy available from the breakdown of 1 mol phosphocreatine, and OGSTON and SMITHIES[223] (1948) (see also DAVIES and OGSTON[97], 1950) have shown that $2{\cdot}0 \times 10^4$ gm cal is available from 1 mol phosphocreatine, so that the breakdown of phosphocreatine provides twice as much energy as was assumed. The discharge also leads to lactic acid formation which is itself a source of energy (NACHMANSOHN *et alii*[219]), and the total electrical energy expended was estimated to be 40 per cent of the energy available from both the breakdown of phosphocreatine and the formation of lactic acid (NACHMANSOHN *et alii*[220]) this becomes 27 per cent if the 1950 figure for energy from phosphocreatine is used. The overall efficiency electrical energy/metabolic energy was estimated at 20 per cent, on the assumption that oxidation would provide about the same amount of energy for the reversal of the two exothermic processes, which are themselves anaerobic. This is the figure quoted in the last column of *Table IX*. The correct figure is likely to be less than this, probably 10–15 per cent.

Bioelectric potentials as a maintained source of electrical energy— Direct evidence of an output of electrical energy by living tissues has been

Table IX. Production of Maintained Electric Power by Various Living Tissues

Material	*Temperature °C*	*Approximate pd across membrane mv*	*Greatest observed current density in active membrane ma/sq cm*	*Greatest observed electrical power output of active membrane μw/sq cm*	*Greatest observed electrical power output of active membrane (μw/mg dry wt)*	*Electrical power/ Metabolic power (Author's figure) per cent*
Halicystis *cell*[28]	25–30	70	0·01 *ma/cell*	0·7 *μw/cell*	—	—
Isolated frog skin[116]	20	—	0·015	2·5	0·5	5–10
Isolated frog skin[256]	25	40	0 02	0 8	0·15	1– 2·5
Isolated frog skin[195]	25	34	0·02	0·9	0·04	1– 2
Dog stomach[236]	39	60	0·15	8·9	0·34	2– 4·5
Isolated frog gastric mucosa[59]	25	40	0·04	1·6	0·26	2
Electric organs[73] *of* Torpedo occidentalis*	—	100	120	12,000	20,000	20**

* Entries refer to instantaneous conditions at peak power of discharge. The average power over a train of discharges is *c* one quarter of the peak power.

** This is electrical energy/metabolic energy per impulse for the main electric organ of *Electrophorus electricus*[220]

obtained mainly from experiments with frog skin, gastric mucosa and electric organs of fishes. *Table IX* is an extended form of Table I of a previous paper[89]. All results in it refer to experiments in which the two sides of the membranes were connected electrically through a circuit of fairly low resistance, and the current densities cited maintained without appreciable polarization. The power produced was limited by the external resistance, which differed in each case, and a higher power would be expected with zero external resistance.

Since the thickness and resistance of the various membranes were not the same, the electrical power output per mg dry wt is not the same fraction of the electrical power output per sq cm for them. STAPP[256] (1941) and LUND and STAPP[195] (1947) used *Rana catesbiana* whose skin is thicker than that of *Rana temporaria* used by FRANCIS[116]. Dog gastric mucosa (REHM[236], 1943) is several times as thick as frog gastric mucosa (CRANE *et alii*[89]) ; moreover the stomach muscle layer was removed from the frog gastric mucosa.

Except for the electric organ, the electrical power output, whether measured per sq cm or per mg dry wt, does not differ by much more than a factor of ten in the membranes listed in *Table IX*. The electric organ is clearly in a class apart as far as electrical power output is concerned, and this is not because the pd across the electrically active membrane is greater, but because the structure is such that much larger current densities are possible. The average electrical power output per mg dry wt during a train of discharges is at least 10^4 times as great as for other structures. The discharges cannot however be continued without periods of rest.

The efficiency, calculated as the fraction : electrical power output divided by the rate at which metabolic energy is liberated, is given in the last column of *Table IX*. These figures represent only the order of magnitude ; most of the calculations by which they were obtained are open to criticism. The average of the three estimations for isolated skin is about 3 per cent, and the average for isolated frog gastric mucosa and dog stomach is about the same. The value for the electric organ (20 per cent) is considerably greater than this ; however it is shown on p 125 that this figure is too high, and that the efficiency is unlikely to be much greater than 10 per cent. LUND and STAPP[195] have discussed the coulomb efficiency of frog skin *i.e.* the number of coulombs generated by frog skin per coulomb equivalent of oxygen consumed. The coulomb efficiency is not a function of energies, but of charges carried by ions.

It must be borne in mind that the structure of electric tissue is especially suitable for the expenditure of electrical energy ; the

electrical power output in frog skin and gastric mucosa was more limited by the resistances in the circuit. We can obtain some estimate of the maximum electrical power output which might be obtained from isolated frog gastric mucosa if the external resistance in the circuit could be reduced to zero, by considering *Figure 7*. This gives the pd current relation for a sheet of frog gastric mucosa both when its own pd was the source of power ($\times$ — $\times$) and when an external source of power was included in the circuit (•—•)[89]. The experimental points, which are typical of those obtained in many experiments, lie about the same straight line whether the gastric mucosa provided its own source of power or not. Now when sheets of frog gastric mucosa were immersed in a beaker of nutrient saline solution at 25·0°C the pd across them, measured between two points close to them, was maintained without polarization at about one fifth of the open-circuit value. These observations suggest that if the resistance external to the mucosa could be reduced to zero, the electrical power developed would be the open circuit pd across mucosa $\times$ current through mucosa when the pd across it is zero. This is about five times the greatest observed electrical power output (for further discussion see CRANE *et alii*[89]).

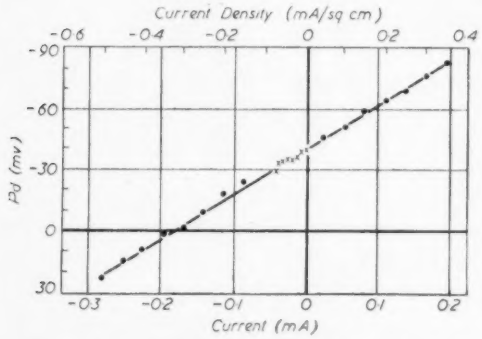


Figure 7. Relation between pd across a frog gastric mucosa and current through it: $\times$ — $\times$ with mucosa as only source of power, • — • with external power source

These results imply that about 10 per cent of the metabolic energy of frog gastric mucosa can be continuously expended as electrical energy. Similar arguments applied to the results for frog skin suggest that this tissue is also capable of expending as electrical power a percentage of its metabolic power considerably greater than that given in the last column of *Table X*. It seems likely therefore, that the efficiency electrical power/metabolic power is similar in frog skin, gastric mucosa and electric organ, in spite of the much greater electrical power output of the electric organ and its different mode of operation.

Utilization of electrical energy provided directly — The capacity of certain tissues to convert an appreciable fraction of the energy available from metabolism into electrical energy raises the question as to whether the expenditure of this electrical energy is intimately bound up with their

function. There is also an auxiliary question which it is convenient to consider first : can such tissues use energy supplied as electrical energy for their normal endergonic functions ?

It has been known for some time that electric current can affect the curvature of growing shoots and roots of plants. WENT and THIMANN[277] (1937) and EVERS[107] (1947) have discussed experiments carried out at the end of the last century and more recent work including that by BRAUMER and BÜNNING[35] (1930), HARTMANN[137] (1931), AMLONG[5] (1933), and KOCH[173] (1934). It is of interest here because the curvature is determined by the distribution of auxins, and the potential gradient produced by an applied electric current might be expected to influence the movement of auxins, since they are anions. There is a considerable amount of evidence that this is so, but Went and Thimann concluded their discussion with the statement 'that the analysis of electrotropic curvatures, so far as it has been carried, is suggestive rather than convincing'. SCHRANK[254] (1945) came to the conclusion that auxins do not play a primary part in determining the pattern of growth.

There is evidence that the rate of secretion of hydrochloric acid by gastric mucosa can be increased or decreased by an applied electric current. An electric current passed through either the stomach of live dogs (REHM[239], 1945) or isolated frog or toad gastric mucosa (CRANE and DAVIES[84], CRANE, DAVIES and LONGMUIR[90]) resulted in an alteration of the rate of acid secretion. The effect was found to be a polar one, currents in such a direction as to enhance the natural potential difference (enhancing currents) increasing the rate of acid secretion, and currents in the opposite direction (opposing currents) decreasing it. With enhancing currents of 1 ma/sq cm the rate of acid secretion of frog gastric mucosa was often increased by more than 100 per cent ; opposing currents of the same magnitude usually decreased it by 100 per cent *i.e.* abolished it. With dog stomach an enhancing current of 1 ma/sq cm produced an increase in the rate of acid secretion only of the order of 30 per cent, whereas the same opposing current produced a decrease more than twice this amount. The change in the rate of acid secretion was roughly proportional to the current density (see *Figure 8*[239]).

With both dog stomach and amphibian gastric mucosa the rate of acid secretion approached its original value when the current was switched off *i.e.* the alteration in the rate of secretion depended upon the immediate supply of electrical energy.

The results just quoted refer to gastric mucosa already secreting acid ; different effects were obtained if there were no acid secretion when the current was switched on. REHM[239] found no initiation of

acid secretion by electric current, up to 7 ma/sq cm, in any of sixteen cases with live dogs. CRANE *et alii*[90] found that in five out of ten cases with secretive isolated frog gastric mucosa, enhancing current up to 1 ma/sq cm did initiate acid secretion, but not in the other five. However 'spontaneous' secretion of acid occurs in isolated frog gastric mucosa, but not in the stomachs of live dogs, so that a difference in behaviour with respect to the initiation of acid secretion is not surprising.

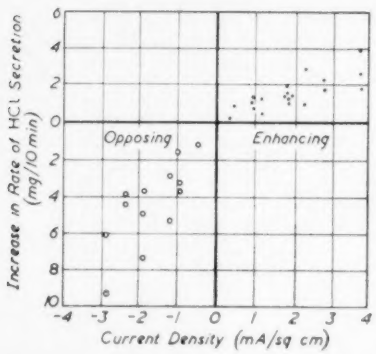


Figure 8. Change in rate of HCl *secretion by dog stomach when an electric current is passed through it. (Enhancing current* ● *is current from serosa to mucosa enhancing natural pd ; opposing current* o *is current from mucosa to serosa, opposing natural pd)*

It is important to stress the specificity of the effect of electric current on the rate of acid secretion. Many experiments were carried out with a variety of other live and dead membranes, and on no occasion was the secretion of any acid detected when a current was passed through them. No evidence was obtained of alkali secretion due to the passage of a current through gastric mucosae or any of the other membranes used.

Too little is known about the processes involved to make an accurate estimate of the efficiency of the utilization of electrical energy for secretion of acid, but attempts have been made with the data available. REHM[239] calculated that the power used in overcoming the emf of dog stomach when an enhancing current was passed through it was less than that used in the production of the extra HCl secreted during the passage of the current. This indicated 'either that the current acts, at least in part, as a trigger mechanism or that some other process furnishes part of the energy necessary for the production of HCl'.

CRANE and DAVIES[84, 86] attempted to calculate the energy relations for isolated frog gastric mucosa. Enhancing current (0·4 ma/sq cm) passed through isolated frog gastric mucosa secreting HCl increased the Q_{HCl} by an average of 1·0. Since the energy required for the secretion of 1 gm mol HCl is probably about $1{\cdot}0 \times 10^4$ cal, the additional Q_{HCl} of 1·0 requires an extra power of 0·5 watt/mg dry wt. According to the electron cycle mechanism of HCl production, an electron is transferred across a Fe^{++}, Fe^{+++} system for each H^+ and Cl^- ion secreted (CRANE *et alii*[89]). This process could be expedited by the pd maintained across the secretory units when an enhancing current is passed through it. For a given mucosa the Q_{HCl} is proportional to

the rate of transport of H^+ ions (and of electrons). Assuming that this rate is proportional to the total pd across the units, the calculated increase in Q_{HCl} agrees with the experimentally observed value. The power required to maintain the enhanced pd is at least that needed to overcome the resistance of the mucosa (Joule effect), which is 4·7 watt/mg dry wt for the average mucosa secreting HCl, plus that needed to overcome the emf of the mucosa, 0·84 watt/mg dry wt. The average overall efficiency of the utilization of directly supplied electrical energy for secretory work would therefore be 0·5/(4·7+0·84) *i.e.* about 10 per cent. REHM[239] also pointed out 'that the total electrical energy delivered to the stomach was greater than the amount of energy needed for the formation of HCl'; the Joule energy was not included in his calculations quoted above.

Bioelectric potentials as a source of energy in endergonic processes— We must now revert to the first question raised on p 127 : is the expenditure of electrical energy closely related to the function of some tissues? This is obviously the case for the electric organ, but in general much less information is available about endergonic than exergonic mechanisms in living tissues, and less is known about the relations between bioelectric potentials and endergonic than exergonic processes.

Suggestions that electrical energy is used for endergonic processes have not been wanting. As long ago as 1859 BRUCKE[39] suggested that the electromotive forces in gastric mucosa may enable it to send an acid fluid towards the secretory surface and a basic fluid in the opposite direction. Since then many workers (*e.g.* ENGELMANN[105], 1872 ; BOENHEIM[32], 1930 ; KELLER[164], 1932) have suggested that glands produce their secretions by electroendosmosis (see also MUDD[214], 1926). Results obtained with acid-secreting gastric mucosa support the view that electrical energy plays a part in the secretion of HCl. The technical problems associated with the simultaneous determination of Q_{O_2}, Q_{HCl} and pd are considerable, but some measurements have been carried out (TERNER[261], 1949).

An attempt has been made to correlate the energies of the exergonic and endergonic processes leading to the secretion of HCl and to find out how the electrical energy output which gastric mucosa can maintain might be involved (CRANE and DAVIES[85, 86]). The energy needed to produce the H^+ and Cl^- ions, together with the water of the secretion, was estimated to be about $1{\cdot}0 \times 10^4$ cal/gm mol HCl. The free energy from glucose oxidation which is available for secretion is unknown. However, estimates which have been made suggest that the thermodynamic efficiency of HCl production calculated on the basis of the Q_{O_2} increase during secretion is very high—at least 50 per cent and possibly more than 100 per cent. It may be, however,

that part of the energy from unstimulated respiration can be utilized for acid secretion. Now since the pd across gastric mucosa decreases and its resistance increases when acid secretion takes place, the electric power available externally (see p 127) decreases at the onset of secretion. It seems possible that some of the electric power previously available externally is utilized within the mucosa during acid secretion, and thus contributes to the power required to maintain this process. Considering average isolated frog gastric mucosa, a Q_{HCl} of 2·0 requires a power of 1·0 watt/mg dry wt. The average increase in Q_{O_2} during acid secretion would provide 1·8 or 0·93 watt/mg dry wt according to two alternative estimates, and the electric power which might be available within the mucosa is about 0·3 watt/mg dry wt. Even if this energy is included. the efficiency is remarkably high.

Many more questions have been raised in this review than have been answered. Moreover the solution of one problem has often raised many more. Clearly much work is needed, particularly with techniques designed so that controlled experiments can be performed with simple biological systems.

This review shows that the range of experiments on bioelectric potentials is so wide with regard to material, technique and method of approach, and that it impinges on so many fields, that it is extremely difficult for a group of workers in one laboratory to acquire sufficient knowledge and experience to assess results obtained in all the others where bioelectric potentials are being studied. It seems probable that the collaboration of a number of workers from different fields, together with a wider appreciation of the special problems and potentialities of the different materials and techniques in use, could lead to rapid advances in our understanding of the processes by which bioelectric potentials are maintained, and of their function.

I wish to thank Dr R. E. Davies for the many helpful suggestions which he has contributed during the preparation of this review.

REFERENCES

[1] ABRAMSON, H. A. *Elektrokinetic phenomena* New York, 1934
[2] — MOYER, L. S. and GORIN, M. H. *Electrophoresis of proteins* New York, 1942
[3] ADAIR, G. S. and GOODMAN, E. N. *J. Physiol.* 87 (1936) 35P
[4] ALTMANN, M. *J. comp. Psychol.* 31 (1941) 481
[5] AMLONG, H. U. *Planta* 21 (1933) 211
[6] ANON. *Dell' uso e dell' attività dell' arco conduttore nelle contrazione dei muscoli* Bologna, 1794
[7] ASHLEY-MONTAGUE, M. F. *Quart. Rev. Biol.* 15 (1940) 211
[8] AUGER, D. and FESSARD, A. *Ann. Physiol. Physicochim. biol.* 15 (1939) 261
[9] BALDWIN, E. *Dynamic aspects of biochemistry* Cambridge, 1947
[10] BARNES, T. C. *Amer. J. Surg.* 69 (1945) 82
[11] — *Anat. Rec.* 94 (1946) 37
[12] — and BEUTNER, R. *Science* 104 (1946) 569
[13] BARRON, E. S. G. *Physiol. Rev.* 19 (1939) 184
[14] — *Ann. Rev. Biochem.* 10 (1941) 1

[15] BARTON, D. S. *Yale J. Biol. Med.* 12 (1940) 335
[16] BARTH, L. G. *Physiol. Zool.* 7 (1934) 365
[17] BERNSTEIN, J. *Elektrobiologie* Braunschweig, 1912
[18] BERTHOLON, M. *De l'électricité des végétaux* Paris, 1783
[19] BETHE, A. *Handbuch der normalen und pathologischen Physiologie* Berlin, 1926–1932
[20] BEUTNER, R. *Die Entstehung elektrischer Strome in lebenden Geweben* Stuttgart, 1920
[21] — *Physical chemistry of living tissues and life processes as studied by artificial imitation of their single phases* Baltimore, 1933
[22] — *Bioelectricity*, in *Medical physics* Chicago, 1944
[23] BIEDERMANN, W. *Elektrophysiologie* Jena, 1895
[24] BISHOP, G. H. *J. cell. comp. Physiol.* 1 (1932) 177
[25] BLACK, V. S. *Canad. J. Res.* 20 (1942) 209
[26] — *Rev. Canad. Biol.* 5 (1946) 311
[27] BLINKS, L. R. *J. gen. Physiol.* 13 (1929) 223
[28] — *Cold Spr. Harb. Sym. quant. Biol.* 1 (1933) 127
[29] — *ibid* 8 (1940) 204
[30] — DARSIE, M. L. and SKOW, R. K. *J. gen. Physiol.* 22 (1938) 255
[31] — and PICKETT, M. J. *ibid* 24 (1940) 33
[32] BOENHEIM, F. *Z. ges. exp. Med.* 71 (1930) 108
[33] BOLING, J. L. and BURR, H. S. *Anat. Rec. Suppl.* 79 (1941) 9
[34] BOURDILLON, R. B. *J. Physiol.* 97 (1939) 138
[35] BRAUNER, L. and BÜNNING, E. *Ber. bot. Ges.* 48 (1930) 470
[36] BREDER, C. M., JR. and SPRINGER, S. *Zoologica* 25 (1940) 431
[37] BRIGGS, L. J., CAMPBELL, A. B., HEALD, R. H. and FLINT, L. H. *Bull. U.S. Dep. Agric.* No. 1379 (1926)
[38] British Museum (Natural History) *List of British vertebrates* London, 1935
[39] BRÜCKE, E. *S.B. Akad. Wiss. Wien* 37 (1859) 131
[40] BUCHTHAL, F. and PÉTERFI, T. *Protoplasma* 27 (1937) 473
[41] BUHMANN, A. *Biochem. Z.* 165 (1925) 122
[42] BULL, H. B. *Physical biochemistry* New York, 1943
[43] — and MOYER, L. S. *J. phys. Chem.* 40 (1936) 9
[44] BURGE, W. E. and BURGE, E. L. *Anesth. & Analges.* 19 (1940) 201
[45] — — *Amer. J. Physiol.* 133 (1941) 231
[46] BURR, H. S. *Proc. nat. Acad. Sci., Wash.* 27 (1941) 276
[46A] — *Yale J. Biol. Med.* 14 (1942) 581
[47] — *ibid* 15 (1943) 823
[48] — *ibid* 16 (1944) 353
[49] — *ibid* 17 (1945) 465
[50] — and BULLOCK, T. H. *ibid* 14 (1941) 51
[51] — HARVEY, S. C. and TAFFEL, M. *ibid* 11 (1938) 103
[52] — HILL, R. T. and ALLEN, E. *Proc. Soc. exp. Biol., N.Y.* 33 (1935) 109
[53] — and HAMMETT, F. S. *Growth* 3 (1939) 211
[54] — and HOVLAND, C. I. *Yale J. Biol. Med.* 9 (1937a) 247
[55] — — *ibid* 9 (1937b) 541
[56] — and MUSSELMAN, L. K. *ibid* 9 (1936) 155
[57] — — *Amer. J. Obstet. Gynaec.* 35 (1938) 743
[58] — — BARTON, D. S. and KELLY, N. B. *Science* 86 (1937a) 312
[59] — — — — *Yale J. Biol. Med.* 10 (1937b) 155
[60] — and NORTHROP, F. S. C. *Quart. Rev. Biol.* 10 (1935) 322
[61] — — *Proc. nat. Acad. Sci., Wash.* 25 (1939) 284
[62] — SMITH, G. M. and STRONG, L. C. *Amer. J. Cancer* 32 (1938) 240
[63] — STRONG, L. C. and SMITH, G. M. *Yale J. Biol. Med.* 10 (1938) 539
[64] — TAFFEL, M. and HARVEY, S. C. *ibid* 12 (1940) 483
[65] BURROWS, H., IBALL, J. and ROE, E. M. F. *Brit. J. exp. Path.* 23 (1942) 253
[66] — ROE, E. M. F. and SCHOBER, B. *Cancer Res.* 5 (1945) 524
[67] CAVALLO, T. *Complete Treatise on Electricity* 4th ed. London, 1795
[68] CAVENDISH, H. *Phil. Trans* 66 (1776) 196
[69] CHAGAS, C. *Anais. Acad. Brasil de Ciencias* 19 (1947) 113
[70] CHAPHEAU, M. *Soc. Biol. Paris* 110 (1932) 97
[71] CHOLODNY, N. G. and SANKEWITSCH, E. C. *Plant Physiol.* 12 (1937) 385
[72] COATES, C. W. and COX, R. T. *Trans. N.Y. Acad. Sci. Ser. II* 1 (1939) 110
[73] — — *Zoologica, N.Y.* 27 (1942) 25
[74] — — ROSENBLITH, W. A. and BROWN, M. V. *ibid* 25 (1940) 249
[75] COPE, O., COHN, W. E. and BRENIZER, A. G., JR *J. clin. Invest.* 22 (1943) 103
[76] CORYELL, C. D. *Science* 92 (1940) 380

77 Cox, R. T. *Amer. J. Phys.* 11 (1943) 13
78 — and Breder, C. M., Jr *Zoologica* 28 (1943) 45
79 — and Coates, C. W. *ibid* 23 (1938) 203
80 — — *ibid* 27 (1942) 27
81 — — *ibid* 28 (1943) 45
82 — — and Brown, M. V. *J. gen. Physiol.* 28 (1945) 187
83 Crane, E. E. Unpublished, 1944
84 — and Davies, R. E. *XVII Internat. physiol. Congr. Abstr. Communications* (1947) 321
85 — — *Biochem. J.* 43 (1948) xlii
86 — — *ibid* 43 (1948) xlii
87 — — *Trans. Farad. Soc.* In press, 1950
88 — — and Longmuir, N. M. *Biochem. J.* 40 (1946) xxxvi
89 — — — *ibid* 43 (1948a) 321
90 — — — *ibid* 43 (1948b) 336
91 Craxford, S. R., Gatty, O. and Teorell, T. *Phil. Mag.* 25 (1938) 1061
92 Curtis, H. J. and Cole, K. S. *J. cell. comp. Physiol.* 19 (1942) 135
93 Damon, E. B. *J. gen. Physiol.* 15 (1932) 525
94 Davies, R. E. *Biochem. J.* 42 (1948a) 609
95 — *The mechanism of hydrochloric acid production by the stomach* Thesis, University of Sheffield, 1948b
96 — Unpublished, 1949
97 — and Ogston, A. G. *Biochem. J.* 46 (1950) 324
98 Davson, H., Duke-Elder, W. S. and Maurice, D. M. *Proc. Physiol. Soc.* 107 (1948) 32P
99 Dean, R. B. and Gatty, O. *Trans. Farad. Soc.* 33 (1937) 1040
100 Delrue, G. *Etude de la sécrétion acide de l'estomac* Thèse Universite Catholique de Louvain, 1935
101 Donnan, F. G. *Z. Elektrochem.* 17 (1911) 572
102 Donne, A. *Ann. Chim. (Phys.)* 57 (1834) 398
103 Du Bois-Reymond, E. *Untersuchungen über tierische Elektrizität* Berlin, 1848
104 Emerson, R. *J. gen. Physiol.* 13 (1929) 153
105 Engelmann, T. W. *Pflüg. Arch. ges. Physiol.* 6 (1872) 97
106 Ettisch, G. and Péterfi, T. *ibid* 208 (1925) 454
107 Evers, J. in Lund, E. J. *Bioelectric fields and growth* Texas, 1947
108 — and Lund, E. J. *ibid* Texas, 1947
109 Ewart, J. C. *Phil. Trans. B.* 179 (1888) 399, 539
110 — *ibid* 183 (1892) 389
111 Faraday, M. *Experimental researches in electricity* London, 1839
112 Fessard, A. *Ann. N.Y. Acad. Sci.* 47 (1946) 501
113 Finch, G., Yerkes, R. M. and Elder, J. H. *Proc. Soc. exp. Biol. N.Y.* 37 (1937) 560
114 Flexner, L. B. and Stiehler, R. D. *J. biol. Chem.* 126 (1938) 619
115 Floyd, W. F. and Keele, C. A. *Trans. Farad. Soc.* 33 (1937) 1046
116 Francis, W. L. *Nature, Lond.* 131 (1933) 805
117 — *J. exp. Biol.* 11 (1934) 35
118 — and Gatty, O. *ibid* 15 (1938) 132
119 — and Pumphrey, R. J. *ibid* 10 (1933) 379
120 Friedenwald, J. S. and Stiehler, R. D. *v. Graefes Arch. Ophthal.* 20 (1938) 761
121 Fritsch, G. *Die elektrische fische, I Malapterurus* (1887), *II Die Torpedineen* Leipzig, 1890
122 Fruton, J. S., Ball, E. G., Bergmann, M., Kalckar, H. M., Meyerhof, O. and Smythe, C. V. *Ann. N.Y. Acad. Sci.* 45 (1944) 357
123 Fulton, J. H. and Cushing, H. *Ann. Sci.* 1 (1936) 239
124 Glasstone, S. *Textbook of physical chemistry* London, 1940
125 Goddard, D. R., in Höber, R., *Physical chemistry of cells and tissues* London, 1946
126 Goodman, E. N. *Surg. Gynec. Obstet.* 75 (1942) 583
127 Gotch, F. *Phil. Trans. B.* 178 (1887) 487
128 — *ibid* 179 (1888) 329
129 — and Burch, G. J. *ibid* 187 (1896) 347
130 Graham, J. and Gerard, R. W. *J. cell. comp. Physiol.* 28 (1946) 99
131 Green, D. E. *Mechanisms of biological oxidation* Cambridge, 1941
132 Greenberg, D. M., Aird, R. B., Boelter, M. D., Campbell, W. W., Cohn, W. E. and Murayama, M. M. *Amer. J. Physiol.* 140 (1943) 47
133 Grundfest, H. *Ann. Rev. Physiol.* 9 (1947) 477
134 Guttman, R. *XVII Internat. physiol. Congr. Abstr. Communications* (1947) 189
135 — and Cole, K. S. *Proc. Soc. exp. Biol. N.Y.* 48 (1941) 293
136 Harman, P. J. *Yale J. Biol. Med.* 14 (1941) 189

137 HARTMANN, H. *Beitr. Biol. Pfl.* 19 (1931) 287
138 HEILBRUNN, L. V. *An outline of general physiology* 2nd ed. Philadelphia and London, 1947
139 HENLY, W. Letter to CANTON, W., undated, 1776 (or possibly early 1777)
140 HINSHELWOOD, C. W. *J. chem. Soc.* (1947) 694
141 HITCHCOCK, D. T., in HÖBER, R., *Physical chemistry of cells and tissues* London, 1946
142 HÖBER, R. *ibid*
143 HODGKIN, A. L. *J. Physiol.* 106 (1947) 319
144 — and HUXLEY, A. F. *Nature, Lond.* 144 (1939) 710
145 — — *J. Physiol.* 104 (1945) 176
146 — — *Nature, Lond.* 158 (1946) 376
147 — — *J. Physiol.* 106 (1947) 341
148 — and KATZ, B. *ibid* 108 (1949) 37
149 HOLLANDER, F. *Gastroenterology* 1 (1943) 401
150 HOLM JENSEN, I. *Kgl. Dansk. Vid. Selsk. Biol. Medd.* 20 No. 11 (1948)
151 HOYT, R. C. *J. cell. comp. Physiol.* 29 (1947) 109, 131
152 HUF, E. *Pflüg. Arch. ges. Physiol.* 232 (1933) 559
153 — *ibid* 235 (1935) 655 ; 236 (1935) 1
154 — *Biochem. Z.* 288 (1936a) 116
155 — *Pflüg. Arch. ges. Physiol.* 237 (1936b) 143 ; 238 (1936b) 97
156 — *ibid* 240 (1938) 572
157 — *Fed. Proc.* 8 (1949) 78
158 HUNTER, J. *Phil. Trans.* 63 (1773) 481
159 — *ibid* 65 (1775) 395
160 JØRGENSEN, C. B., LEVI, H. and USSING, H. H. *Acta physiol. Scand.* 12 (1947) 350
161 JOSEPH, N. R., REED, C. I., KAPLAN, E. and STECK, I. E. *Amer. J. Physiol.* 157 (1949) 63
162 — — STECK, I. E., FOLK, F. and KAPLAN, E. *ibid* 153 (1948) 364
163 KATZ, B. *Proc. roy. Soc.B* 135 (1948) 506
164 KELLER, R. *Die Elektrizität in der Zelle.* 3rd ed. Mährisch-Ostrau, 1932
165 KEYNES, R. D. *Proc. physiol. Soc.* 107 (1948) 35P
166 — *ibid* 108 (1949) 9P
167 KEYS, A. *Proc. roy. Soc.B* 112 (1933) 184
168 — *Trans. Farad. Soc.* 33 (1937) 972
169 KHALIL, F. *Z. vergl. Physiol.* 25 (1937) 256
170 KINSEY, V. E. *J. Physiol.* 32 (1949) 329
171 — GRANT, W. M., KOGAN, D. A., LIVINGOOD, J. J. and CURTIS, B. R. *Arch. Opthal., Chicago* 27 (1942) 1126
172 KLAUS, K. *Biochem. Z.* 185 (1927) 3
173 KOCH, K. *Planta* 22 (1934) 1
174 KOMAGATA, S. *Res. Electrotech. Lab., Tokyo* No. 362 (1934)
175 KORR, I. M. *Cold Spr. Harb. Sym. quant. Biol.* 7 (1939) 74
176 KRAJUCHIN, B. V., BODROVA, M. B. and MOGILA, M. T. *J. Méd. Ukrain* 8 (1938) 676, 682, 689
177 KREBS, H. A. Unpublished experiments quoted in KREBS, H. A. and JOHNSON, W. A. *Tabu. biol, Berl.* 19 (1948) part 3
178 — and HENSELEIT, K. *Hoppe-Seyl. Z.* 210 (1932) 33
179 — and JOHNSON, W. A. *Tabu. biol., Berl.* 19 (1948) part 3
180 — and ROSENHAGEN, H. *Z. Neurol. Psych.* 134 (1931) 643
181 KROGH, A. *Trans. Farad. Soc.* 33 (1937) 912
182 — *Proc. roy. Soc.B* 133 (1946) 140
183 KUMANOMIDO, S. *Biochem. Z.* 193 (1928) 315
184 KÜMMEL, K. *Planta* 9 (1929) 564
185 KURZROK, R. *The endocrines in obstetrics and gynaecology* London, 1937
186 LANGMAN, L. and BURR, H. S. *Amer. J. Obstet. Gynec.* 44 (1942) 223
187 LIPMANN, F. *Advances in Enzymology* 1 (1941) 99
188 LIPMANN, G. *Ann. Chim. Phys.* 5 (1875) 494
189 LOEB, J. *Proteins and the theory of colloidal behaviour* London, 1922
190 LUND, E. J. *J. exp. Zool.* 51 (1928) 265, 291, 327
191 — *Bioelectric Fields and Growth* Texas, 1947
192 — and GUNTER, G., in *ibid* 1947
193 — MAHAN, R. I. and HANSZEN, A. H. *Proc. Soc. exp. Biol., N.Y.* 60 (1945) 326
194 — — — in *Bioelectric fields and growth* Texas, 1947
195 — and STAPP, P., in *ibid* 1947
196 LUNDEGÅRDH, H. *Ark. Bot.* 32 (1945) 1

REFERENCES

[197] LUNDEGÅRDH, H. *Nature, Lond.* 157 (1946) 575
[198] — *Discussions of the Faraday Society* No. 3 (1948) 139
[199] — *Ann. roy. agric. Coll. Sweden* 16 (1949) 372
[200] MATTHEWS, B. H. C. *J. Instn. Elect. Engrs I* 95 (1948) 397
[201] MAXWELL, J. C. *The electrical researches of the Honourable Henry Cavendish, 1771–1781* Cambridge, 1879
[202] MEYER, K. H. *Trans. Farad. Soc.* 33 (1937) 1049
[203] — and BERNFELD, P. *J. gen. Physiol.* 29 (1946) 353
[204] MEYERHOF, O., LOHMANN, K. and MEIER, R. *Biochem. Z.* 157 (1925) 459
[205] MICHAELIS, L. *J. gen. Physiol.* 8 (1925) 33
[206] — *Colloid Symposium Monographs* 5 (1928) 135
[207] MILLER, J. R. and DENT, R. F. *J. lab. clin. Med.* 28 (1943) 168
[208] MISLOWITZER, E. and SILVER, S. *Biochem. Z.* 256 (1932) 432
[209] — — and ROTHSCHILD, M. *ibid.* 256 (1932) 444
[210] MOND, R. *Pflüg. Arch. ges. Physiol.* 215 (1927) 468
[211] MOULLIN, E. B. *Spontaneous fluctuations of voltage due to Brownian Motions of electricity, shot effect and kindred phenomena* Oxford, 1939
[212] MUDD, S. *J. gen. Physiol.* 7 (1925) 389
[213] — *ibid* 9 (1925) 73
[214] — *ibid* 9 (1926) 361
[215] NACHMANSOHN, D., COATES, C. W. and COX, R. T. *ibid* 25 (1941) 75
[216] — — and ROTHENBERG, M. A. *J. biol. Chem.* 163 (1946) 39
[217] — COX, R. T., COATES, C. W. and MACHADO, A. L. *J. Neurophysiol.* 5 (1942) 499
[218] — — — — *Proc. Soc. exper. Biol. N.Y.* 52 (1943a) 97
[219] — — — — *J. Neurophysiol.* 6 (1943b) 383
[220] — COATES, C. W., ROTHENBERG, M. A. and BROWN, M. V. *J. biol. Chem.* 165 (1946) 223
[221] NICHOLAS, J. S. and CARMOSINO, J. A. *Yale J. Biol. Med.* 17 (1944) 235
[222] OGSTON, A. C. *Physiol. Rev.* 27 (1947) 228
[223] — and SMITHIES, O. *ibid* 28 (1948) 283
[224] OSTERHOUT, W. J. V. *Trans. Farad. Soc.* 33 (1937) 997
[225] — *Cold Spr. Harb. Sym. quant. Biol.* 8 (1940) 51
[226] — *J. gen. Physiol.* 26 (1943) 293
[227] PARMENTER, R. *Amer. J. Physiol.* 126 (1939) 597
[228] PATTERSON, W. B. and STETTEN J. R. DE W. JR. *Science* 109 (1949) 256
[229] PLEDGE, H. T. *Science since 1500* London, 1939
[230] PONDER, E. and MACLEOD, J. *J. gen. Physiol.* 20 (1937) 433
[231] PUMPHREY, R. *Arch. exp. Zellforsch.* 15 (1934a) 261
[232] — *J. exp. Biol.* 9 (1934b) 423
[233] QUIGLEY, J. P., BARCROFT, J., ADAIR, G. S. and GOODMAN, E. N. *Amer. J. Physiol.* 119 (1937) 763
[234] REBOUL, J., DAVIS, H. and FRIEDGOOD, R. B. *ibid* 120 (1937) 724
[235] — FRIEDGOOD, R. B. and DAVIS, H. *ibid* 119 (1937) 387
[236] REHM, W. S. *ibid* 139 (1943) 1
[237] — *ibid* 140 (1944a) 720
[238] — *ibid* 141 (1944b) 537
[239] — *ibid* 144 (1945) 115
[240] — *ibid* 147 (1946) 69
[241] — and ENELOW, A. J. *Gastroenterology* 3 (1944) 306
[242] — and HOKIN, L. E. *Amer. J. Physiol.* 149 (1947) 162
[243] RICE, H. V. and ROSS, R. T. *ibid* 149 (1947) 77
[244] ROBERTSON, R. N. Private communication, 1948
[245] — and WILKINS, M. J. *Nature, Lond.* 161 (1948a) 101
[246] — — *Austr. J. Sci. Res.* 1 (1948b) 17
[247] ROGERS, P. V. *Proc. Soc. exp. Biol., N.Y.* 34 (1936) 644
[248] — *ibid* 37 (1937) 212
[249] — *Endocrinology* 22 (1938) 35
[250] ROSENE, H. F. *Bibliography of continuous bioelectric currents and bioelectric fields in plants and animals* Included in LUND, E. J., 1947
[251] ROTHSCHILD, LORD *J. exp. Biol.* 23 (1947) 267
[252] SACKS, J. *Chem. Rev.* 42 (1948) 411
[253] SARRE, H. *Z. Biol.* 95 (1934) 135
[254] SCHRANK, A. R. *Plant Physiol.* 20 (1945) 133
[255] SNODGRASS, J. M., ROCK, J. and MENKIN, M. F. *Amer. J. Physiol.* 140 (1943) 394

[256] STAPP, P. *Proc. Soc. exp. Biol., N.Y.* 46 (1941) 382
[257] STEINBACH, H. B. *J. cell. comp. Physiol.* 7 (1935) 291
[258] STIEHLER, R. D. and FLEXNER, L. B. *J. biol. Chem.* 126 (1938) 603
[259] SWYNGEDAUW, R. *C. R. Soc. Biol., Paris* 98 (1928) 1431
[260] TEORELL, T. and WERSÄLL, R. *Acta physiol. Scand.* 10 (1945) 243
[261] TERNER, C. Thesis, University of Sheffield, 1949
[262] TITAEV, A. A. *Bull. Biol. Med. exp.* 5 (1938)
[263] UMRATH, K. *Protoplasma* 17 (1933) 258
[264] — *ibid* 31 (1938) 184
[265] USSING, H. H. *Acta physiol. Scand.* 17 (1949) 1
[266] — *Physiol. Rev.* 29 (1949) 127
[267] VAN SLYKE, D. D. *Factors affecting the distribution of electrolytes, water and gases in the animal body* Philadelphia Pa, 1926
[268] VENCHIKOV, A. I. *J. Physiol. U.S.S.R.* 25 (1938a) 478
[269] — *Bull. Biol. Med. exp.* 5 (1938b)
[270] VISSCHER, M. B., FLETCHER, E. S., CARR, C. W., GREGOR, H. P., BUSHEY, M.S. and BARKER D. E. *Amer. J. Physiol.* 142 (1944) 550
[271] VOLTA, A. *Giornale Fisico-Medico* 1792
[272] — Letter to Sir Joseph Banks 1800
[273] WALLS, G. L. *The Vertebrate Eye* Michigan, 1942
[274] WALKER, W. C. *Ann. Sci.* 2 (1937) 84
[275] WALLER, A. D. *On Animal Electricity* 1897
[276] WALSH, J. *Phil. Trans.* 64 (1774) 465
[277] WENT, F. W. and THIMANN, K. V. *Phytohormones* New York, 1937
[278] WIGGLESWORTH, V. B. *Proc. roy. Soc.B* 135 (1948) 430
[279] WOHLGEMUTH and KLOPSTOCK, E. *Biochem. Z.* 175 (1926) 202

5

PHASE-CONTRAST MICROSCOPY IN BIOLOGICAL RESEARCH

A. F. W. Hughes

THE phase-contrast microscope is rapidly becoming a standard instrument in biological laboratories. Although it has become available only in the last two years, its history extends over a longer period. ZERNIKE's original papers[1] on the principle of phase contrast date from 1934; the application to the microscope followed soon afterwards[2].

At this time, BURCH in this country was using a Zernike microscope, but despite considerable efforts failed to convince either biologists or manufacturers of its importance. BRICE[3] states that apparently the Zeiss firm were only slowly persuaded by Zernike himself to take up the invention. Papers by KÖHLER and LOOS[4] and MICHEL[5] on the function and use of the phase-contrast microscope appeared under Zeiss auspices in the early 1940's. In 1942 BURCH and STOCK[6] described a method of preparing and fitting a phase-plate to a standard 2 mm objective.

The present period of general interest in the phase microscope dates from the time when in 1946 MICHEL's memorable film on the spermatogenesis of a grasshopper was brought to this country, and the phase-contrast principle acquired the status of a German technical secret.

The explanation of the phase-contrast principle given by BURCH and STOCK[6] will probably suffice for most biologists who use the instrument. More recent papers on the theory of the instrument are those of MARTIN[7], PAYNE[8] and BERTI[9]. Among technical articles on various aspects of the phase microscope are those by BENNETT[10] *et alii*, TAYLOR[11] and OSTERBURG[12].

PHASE MICROSCOPY IN BIOLOGY

The purpose of the phase microscope is to form a magnified image of a transparent unstained object in which are slight inhomogeneities of refractive index. For the biologist, the obvious application is to living cells and tissues, although its use in biology is not confined to unfixed material. BARER[13] has shown that the phase microscope may advantageously be used on sections in which a specific staining reaction leaves the general features of the tissue unrevealed by the ordinary microscope. PELC[14] examines in this way sections in contact with

autoradiographs, to obviate the risk of displacement of the section during ordinary histological procedures.

Wherever a suspension of cells, or a thin layer of a tissue can be prepared without distortion, a phase microscope can almost certainly be used to study the form and arrangement of cells. Barer has even been able to study, with its aid, details of the nervous system within an intact insect larva. Where, however, the unit is the individual cell rather than the tissue, we are necessarily limited to those cells in which differences of refractive index are to be found in life. Probably most, if not all cells of vertebrate animals are of this kind, but it seems probable that many plant cells are uniformly refractile. Among the spermatocytes of various orthoptera, there is considerable variation in this respect as is shown by BELAR[15].

The assistance which is given by the phase microscope in the study of a transparent object is naturally greatest where the inequalities of refractive index are too small to enable them to be revealed adequately by the ordinary microscope. Cells in the outgrowth of cultures of vertebrate tissues come into this category (ZOLLINGER[16], HUGHES and SWANN[17]). Although the ordinary microscope in the hands of CANTI[18] and LEWIS[19] has been of great value for the study and photographic recording of living cells in culture, the phase microscope reveals much finer detail within the nucleus of such cells than can be seen by any other means.

With the normal positive phase-contrast system, a granule in the object which gives rise to a retardation of $\lambda/4$ or less, when in focus appears in the image as a dark spot on a lighter background. Immediately on either side of the focal point the contrast is reversed. Beyond this point out of focus detail seems not to affect the image, and so it is possible to focus through an object such as a living cell, and bring into view a whole series of entirely separate images through vertical distances of the order of 10 μ or more. The depth of focus of a phase-contrast system at high powers is almost nil. The phase microscope can be used on objects much thicker than is possible with dark ground illumination, provided that nowhere within the object are retardations which greatly exceed $\lambda/4$.

Where the retardation in the object exceeds $\lambda/4$, the contrast of the object detail decreases, and ultimately, beyond the point of phase reversal it is replaced by brightness and glare. In an object where large retardations occur, the study of the finer detail is much impeded. Thus in cells in the outgrowth of tissue cultures, fine detail within the nucleus and the chromosomes during division can readily be seen, provided that the highly refractile lipoid material in the cytoplasm is not too abundant. Where phase reversal is predominant in an object,

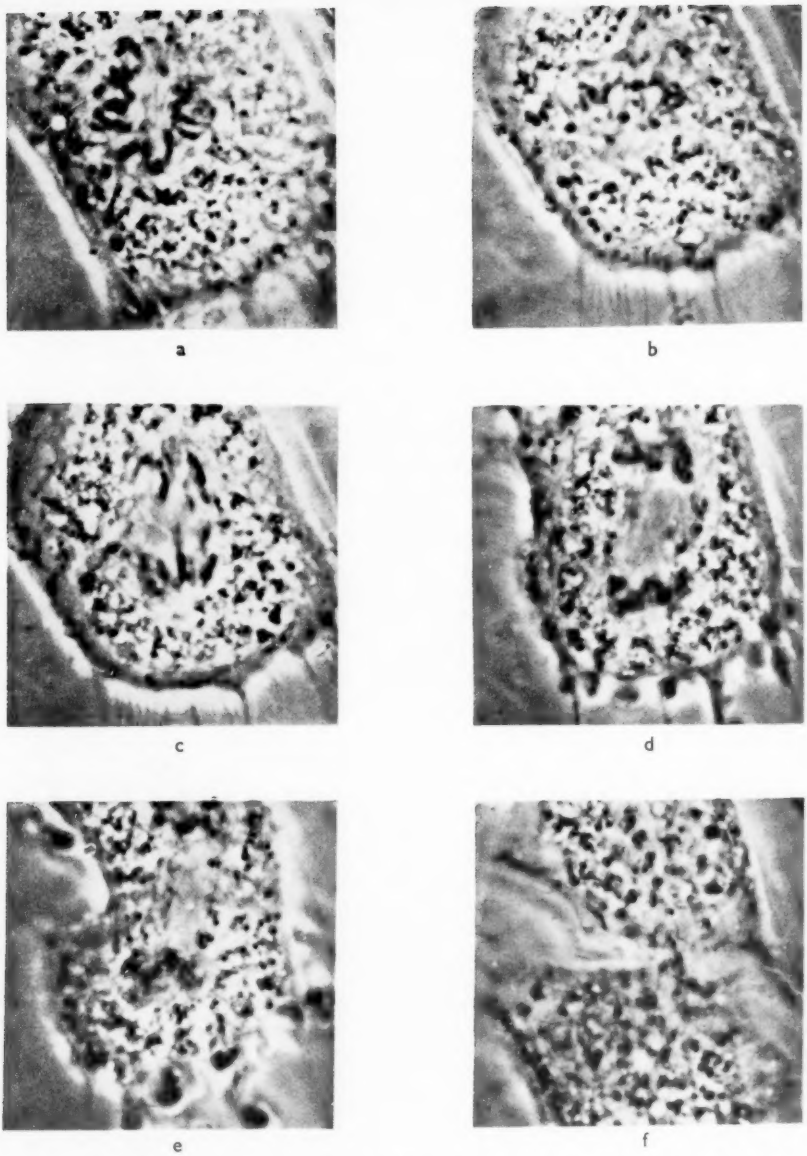


Figure 1. Cell from mouse spleen culture in division (culture by Fell). Phase-contrast × 2,000

a *End of prophase. The nuclear membrane has gone and chromosomes are fully formed*
b *8 minutes later than* a. *Beginning of anaphase.*
c *10 minutes later than* a
d *15 minutes later than* a
e *17 minutes later than* a. *Cleavage of cell in progress*
f *30 minutes later than* a. *Cleavage complete. Daughter nuclei in reconstruction; in them nucleoli and nuclear membrane can be seen*

then the phase microscope has no advantage over the ordinary bright field method.

PHOTOGRAPHY

Preparations which are studied under the phase microscope are often transient, and if living, may be in continual change. Photography is therefore of great importance in recording observations, as shown by the work of RICHARDS[20]. The technique of serial photography on cine film in the study of living cells by phase micrography has been described by HUGHES[21]. For demonstrating fine detail in the cell, the contrast given by the phase microscope needs to be enhanced by the use of a high contrast film of slow speed. A limit is set by the light sensitivity of the living object, but it has been found possible to follow a cell through mitosis and for several hours afterwards, when photographed at intervals of a few seconds on a high contrast 16 mm film. This is illustrated by the series of photo-micrographs in *Figure 1* ; the culture was prepared by Fell.

Since phase-contrast study is generally combined with photomicrography, the contrast in the resulting photograph can be readily varied by choice of the sensitized materials, when a standard phase-contrast objective is used. A system of varying contrast by optical means alone has been described by TAYLOR[22]. In both methods, increase of contrast requires that more light is transmitted through the object.

LITERATURE ON PHASE-MICROSCOPICAL RESEARCH

Since the phase microscope has only become generally available within the last two years, only a few papers on research involving its use have so far appeared. These, however, have indicated some of the lines along which further advances are to be expected.

LUDFORD, SMILES and WELCH[23] have studied cells of a malignant tumour and have compared their appearance by phase-contrast and ultraviolet microscopy. Some of the photographs of this material which these authors have exhibited at demonstrations within the past year must be among the most striking photomicrographs ever produced. AUSTIN and SMILES[24] have studied the fertilization and early development of the rat egg by phase contrast, and have shown that even for cells of the size and thickness of a mammalian ovum the phase microscope can be used to great advantage. Various aspects of mitosis of cells in tissue culture have been studied by HUGHES and SWANN[17], HUGHES and FELL[25], HUGHES and PRESTON[26] both in normal material and under the influence of chemical reagents.

ZOLLINGER[16, 27–29] has published a series of papers on experimen-

tal cytology. He has studied the effect of a variety of treatments on suspension of cells, for instance from the frog kidney and from human tumours. The formation of blisters on the cell membrane, the appearance of mitochondria during cytolysis, and the effect of fixatives on the nucleus are among the topics with which Zollinger has been concerned. OETTLE[30] has examined living human spermatogonia by phase-contrast and concludes that ' structures are revealed which resemble the Golgi apparatus of impregnated cells ', but that these are inconstant in form.

RICHARDS[31] has examined different species of living bacteria in cultures by phase microscopy. He stressed the value of the method in measuring living cells, and gives a number of results. He states that ' the uncertainty of measurements reported is less than the usual figures of 0·2 μ for white light '. Some of the larger bacteriophages have also been seen by phase-contrast, and the lysis of bacterial cells by bacteriophage has been followed by this means by HOFER and RICHARDS[32] and RICHARDS[31].

The apparent increase of resolution given by the phase microscope has been utilized by BARER[33] to observe the elementary bodies of psittacosis and vaccinia virus. These are rendered visible in the phase microscope although strictly speaking they are not resolved.

In a few years' time the literature on the use of the phase microscope will probably be much larger. Of the newer developments in microscopical methods during the last decade, phase-contrast may well prove the most important, because it alone enables living cells to be studied without injury. But where it is used in combination with other methods, we may expect the most fruitful results.

REFERENCES

[1] ZERNIKE, F. *Mon. Not. roy. Astronom. Soc.* 94 (1934) 377
[2] — *Z. Tech. Phys.* 16 (1935) 454
[3] BRICE, A. T. *The Phase Principle in Microscopy* F.I.A.T. Final Report 1059, 1947
[4] KÖHLER, A. and LOOS, W. *Naturwiss.* 29 (1941) 49
[5] MICHEL, K. *Naturwiss.* 29 (1941) 61
[6] BURCH, C. R. and STOCK, J. P. *J. sci. Instrum.* 19 (1942) 71
[7] MARTIN, L. C. *Nature, Lond.* 159 (1947) 827
[8] PAYNE, B. O. *J. sci. Instrum.* 24 (1947) 163
[9] BERTI, L. *Micro. Cimento* 5 (1948) No. 3
[10] BENNETT, A. H., JUPNIK, H., OSTERBERG, H. and RICHARDS, O. W. *Trans. Amer. micr Soc.* 65 (1946) 99
[11] TAYLOR, E. W. *J. R. micr. Soc.* 66 (1946) 1
[12] OSTERBURG, H. *J. opt. Soc. Amer.* 38 (1948) 685
[13] BARER, R. *Quart. J. micr. Sci.* 88 (1947) 491
[14] PELC, S. R. *Nature, Lond.* 160 (1947) 749
[15] BELAR, K. *Protoplasma* 9 (1930) 209
[16] ZOLLINGER, H. V. *Mikroscopie* 3 (1948) 1
[17] HUGHES, A. F. and SWANN, M. M. *J. exp Biol.* 25 (1948) 45
[18] CANTI, R. G. *Arch. exp. Zellforsch* 6 (1928) 86
[19] LEWIS, W. H. *ibid* 23 (1939) 8

REFERENCES

[20] Richards, O. W. *J. biol. photogr. Ass.* 16 (1947) 29
[21] Hughes, A. F. *J. R. micr. Soc.* 69 (1949) 53
[22] Taylor, E. W. *Proc. roy Soc.A.* 190 (1947) 422
[23] Ludford, R. J., Smiles, J. and Welch, F. V. *J. R. micr. Soc.* 68 (1948) 1
[24] Austin, C. R. and Smiles, J. *ibid* 68 (1948) 13
[25] Hughes, A. F. and Fell, H. B. *Quart J. micr. Si.* 90 (1949) 37
[26] — and Preston, M. M. *J. R. micr. Soc.* 69 (1949) 121
[27] Zollinger, H. V. *Amer. J. Pathol.* 24 (1948) 545
[28] — *ibid* 24 (1948) 569
[29] — *ibid* 24 (1948) 797
[30] Oettlé, A. G. *Nature, Lond.* 162 (1948) 76
[31] Richards, O. W. *Stain Technol.* 23 (1948) 55
[32] Hofer, A. W. and Richards, O. W. *Science* 101 (1945) 466
[33] Barer, R. *Nature, Lond.* 162 (1948) 251

6

LOCAL REFRACTOMETRY: THE OBSERVATION OF TRANSPARENT INHOMOGENEOUS MATTER

J. St. L. Philpot

In the examination of biological material, whether *in situ* or isolated, the unspecific distribution of total matter is usually of interest in addition to specific distributions of individual substances. Both sorts of distribution can often be measured optically, the unspecific by local refractometry, and specific ones by local absorptiometry. For instance in visual microscopy colourless specimens are observed by local refractometry and stained ones by local absorptiometry, with different stains revealing different substances. Similarly in the measurement of sedimentation constants, diffusion constants, and electrophoretic mobilities of dissolved substances, local refractometry is used for observing all the major components in a mixture, while local absorptiometry often can be used for observing one or more minor components in the presence of large amounts of less interesting major ones.

The measurements of sedimentation constants *etc* are so arranged that the distribution of substances is inhomogeneous in only one dimension. In microscopy the inhomogeneities are three dimensional, though they are often reduced to two dimensions by using thin sections. This review is a general account of methods used in one, two and three dimensional local refractometry. They are mostly adapted from well known methods for testing lenses and mirrors[1–3]. Local refractometry *etc* will be referred to for brevity merely as refractometry *etc*.

OBSERVATIONS IN THREE DIMENSIONS

The three dimensional case may be dismissed briefly because although it is the most important one in microscopy no quantitative theory of it seems to exist. This is presumably because the easier two dimensional theory still keeps mathematicians fully employed. The three dimensional problem is particularly complicated because light used for observing any one inhomogeneity can only reach it through neighbouring inhomogeneities. A measurement therefore gives not the refractive index but some integral involving, roughly speaking, the refractive index at all points divided by distance from a chosen point. The refractive index is only obtainable from this integral by solving an integral equation. The problem is like that of x-ray crystal analysis

except that the light is focused by lenses and that the object is usually an aperiodic continuum instead of a periodic point lattice.

For qualitative purposes the assumption is usually made that if a thin and not too inhomogeneous system like a living cell is uniformly illuminated, and if a plane through the system, called the object plane, is focused by a lens to give an image plane said to be conjugate with the object plane, then a small refractile particle in the object plane will be represented at the conjugate point in the image plane by a highly localized retardation of phase. This has been proved to be true in the two dimensional case, within limits set by the resolving power of the system, and under favourable conditions it is undoubtedly a good approximation even in the three dimensional case.

The magnitude of the retardation in the image depends roughly on the product of refractive index and thickness of the particle, averaged over all directions within the illuminating beam. The greater the aperture the less will be the effect of particles not in the object plane, *i.e.* the shorter will be the focal depth. Thus a high aperture system cuts a thin optical section and can reduce the problem in favourable cases from a three to a two dimensional one. Three dimensional space perception can still be provided by a binocular eyepiece, which enables a pile of image planes to be viewed stereoscopically, each image plane giving an optical section of its own object plane. In this process the eye presumably accommodates and converges on successive image planes in rapid succession, using the same coordinating reflexes as in normal binocular vision.

In questions of aperture a distinction is necessary between illumination aperture and viewing aperture. The scale of the diffraction pattern in the image of a solitary point object, expressed as distances in the object plane, depends solely on the viewing aperture, but the interaction between images of adjacent objects depends on the coherence of their illumination, which depends in turn on the illumination aperture. Coherence may be defined as the mutual property of two wave motions which determines whether they add by vector addition of amplitudes or by scalar addition of intensities (see Appendix I). The importance of high aperture illumination is not only that it gives more light but that by diminishing the coherence it decreases the interaction between adjacent images *i.e.* it increases the resolving power and decreases the focal depth.

OBSERVATIONS IN ONE AND TWO DIMENSIONS

One and two dimensional refractometry differ from each other only in two points. In one dimensional refractometry the second dimension is available for automatically plotting a graph of the refractive index or

its gradient,[29-33, 35-38] the first dimension being used as abscissa to show position in the object. In two dimensional refractometry both dimensions are needed to show position, and the refractive index can only be shown as contour lines along which it or its gradient is constant. The optical stops are usually rectangular for one dimension and circular for two dimensions, to give a radially symmetrical viewing system. The one dimensional diffraction patterns therefore depend on Bessel functions of half-integral order, such as $(\sin x)/x$ and its integral Si (x), together with the corresponding cosine functions, while the two dimensional patterns depend on Bessel functions of integral order.

When an image of an inhomogeneous illuminated object is formed by a lens, the inhomogeneities deviate the light and the lens corrects these deviations, returning the light approximately to the conjugate image point. This phenomenon may be used in two distinct ways, each the basis of several methods of study. In what may be called afocal methods the object is out of focus so that the deviation can be directly observed, though the deviating point in the object must be determined indirectly. In focal methods the deviation is cancelled by focusing the object, so that the deviating point is obvious but the deviation must be measured indirectly.

The self-plotting one dimensional methods can be described as astigmatic since the object is defocused in the second dimension. They can still be classed as afocal or focal in the first dimension. The varieties of afocal and focal methods will now be enumerated.

AFOCAL METHODS

X-ray crystal analysis— Though not optical, this technique should be mentioned as an interesting example of an afocal method. Here the deviation results from periodic inhomogeneities and varies inversely with their spacing. Their positions are of no interest apart from their spacing, so that the problem of location is obviated. The Abbé spectra[4], observable in a microscope objective when viewing a periodic object come under this heading. With an aperiodic object the first order spectra broaden and the higher orders become faint though still important. Their chief interest for present purposes is in helping to understand some of the methods described below, but care is necessary to avoid attributing to aperiodic objects phenomena peculiar to periodic ones.

Narrow pencil illumination— This form of illumination in microscopy is an afocal method, consisting merely in observing a slightly defocused object with small illumination aperture. The deviated light is observed against a bright coherent background as near to the focus as possible, to reduce the area within which the originating inhomogeneity must

be located by guesswork. Though this method is unsound, as emphasized by ZERNIKE[4], it is the most convenient of all and can give surprisingly good results in skilful hands. It is therefore widely used for observing tissue cultures. The contrast obtainable decreases with illumination aperture while the brightness and resolving power, as effected by coherence, increase. Success thus depends on a careful adjustment of both focus and illumination aperture.

Lamm scale method— For studies in one dimension this was the first refractometric method[5] to be devised for the ultracentrifuge and though laborious it is still possibly the most accurate. A fine scale is photographed through a slightly defocused ultracentrifuge cell, the refractive gradients of which deviate the images of the scale lines. The deviations are measured by comparison with an undeviated image, and the location of the inhomogeneity in the cell is based on the fact that light reaching a deviated image has passed through the cell at the level of the undeviated image. As with most afocal methods, a small aperture is necessary. If the illumination aperture is small, it should be permissible for the viewing aperture to be large and *vice versa*. The former arrangement gives a narrower diffraction pattern and therefore greater resolving power, but in practice the latter is used, presumably to avoid the confusion which arises if the diffraction patterns are too sharp and do not blend with their neighbours. A short focus modification of Lamm's scale method was introduced recently by SVENSSON[6].

Gouy's method— One of the earliest afocal methods, due to GOUY[7], was recently rediscovered on both sides of the Atlantic[8-11]. This method is probably the best for the measurement of diffusion constants under ideal conditions where the gradient of refractive index is related to position by the simple Gaussian formula $\exp(-x^2)$. It is not suitable for the observation of moving boundaries as in sedimentation and electrophoresis. The undeviated light from a slit is focused to a horizontal line image, which is spread out over a considerable height in the form of overlapping diffuse Abbe spectra when the light is deviated by the diffusion cell placed near the lens. By the ' principle of stationary phase '[12] each level in the deviated image receives light mainly from two levels in the diffusion cell, above and below the middle, where the gradients have a value proportional to the deviation in question. At certain levels these two contributions are out of phase and give black interference fringes which move towards the middle as the gradients are diminished by diffusion. From their rate of movement the diffusion constant can be obtained accurately even in dilute solutions. The method has been extended by OGSTON[13] to the analysis of mixtures. By counting the number of fringes and inserting a double stop which converts the instrument into a RAYLEIGH refractometer[14], the total

retardation due to the undiffused solute needed for calculating the diffusion constant can be obtained accurately.

Rayleigh refractometer— In its original form this may be described as a zero-dimensional afocal method, so arranged that interference fringes are deviated proportionately to difference of retardation in two rectangular homogeneous areas. It has been adapted by CLAESSON[15] for continuous measurement of concentration of eluate in a chromatographic column (' frontal analysis '). PHILPOT and COOK[16] transformed it into an astigmatic focal method for automatically plotting retardation against height in an ultracentrifuge cell. This was done as usual by means of a cylindrical lens which focused the cell in one dimension while the deviated fringes in the slit image were focused in the other dimension. The same transformation can be applied[17] to another well known zero-dimensional afocal method which was made into a self-plotting photoelectric method for flowing eluates by Claesson. Here the fluid object is contained in a hollow prism ; light from a vertical slit is deviated sideways proportionately to the refractive index. In the astigmatic form the prism is focused vertically by a cylindrical lens while the slit is still focused horizontally. Rough tests of this with diffusion boundaries seemed quite promising, and the method might be useful for electrophoresis, but for the ultracentrifuge it involves a difficult problem in cell construction and is probably impracticable.

THEORETICAL TREATMENT FOR FOCAL METHODS

As explained above the term focal is used here to describe any method of refractometry in which the plane of observation is conjugate with the object whose retardation is to be studied. The basis of all focal methods has been clearly stated by ZERNIKE[18] in discussing his phase-contrast method which is one of them. The present statement is developed from Zernike's in an attempt to include the practically important but theoretically intractable case of an extended light source and an objective of finite aperture.

Consider an unpolarized forward-travelling beam of light defined by its amplitude F. The definition of F, which for present purposes is a scalar complex amplitude, is given in Appendix II. Let $F_{jk}(r_j,\phi_j,r_k,\phi_k)$ be the value of F at the point (r_k, ϕ_k) on the normal plane surface $z = z_k$, due to a unit oscillation at the point (r_j,ϕ_j) on $z = z_j$, where z,r,ϕ are cylindrical coordinates and $j,k = 0,1,2$, *etc.* Then Huygens' principle can be expressed in the form

$$F_{jk} = \int_0^{r_{Bm}} r_m \int F_{jm} F_{mk}\, d\phi_m\, dr_m = H_m F_{jm} F_{mk} \qquad \ldots . (1)$$

where H_m may be called the 'Huygens' operator', commuting with any quantity which is independent of r_m, and $z = z_m$ is any intermediate surface, while $r_{Bm} = S_{Bm}(\phi_{Bm})$ is any function defined by a stop S_m. The region of integration *i.e.* where light gets through, will be called the bright part of S_m, and written S_{Bm}, and the rest will be called the dark part, S_{Dm}. If the surface z_0 is an incoherent light source of unit surface brightness, the intensity *i.e.* the observed quantity that affects the eye or a photographic plate, is given by

$$\int_0^{r_{Bo}} \int^{\phi_{Bo}} F_{ok} F_{ok} d\phi_o dr_o = IF_{ok} \qquad \ldots\ldots(2)$$

where I may be called the ' intensity operator ', which when acting on any quantity independent of (r_o,ϕ_o), merely converts it to the square of its magnitude. Now let $z = z_1$ be an object surface with a field stop S_1. In practice both S_0 and S_1 are projected images, and the light at z_0 is not completely incoherent since the source is further back and is projected on to either z_0 (Koehler illumination) or z_1 (critical illumination)[47] ; but these complications will be neglected for present purposes.

On the object surface z_1 let there be an object represented by $O(r_1,\phi_1)$ a variable complex number whose magnitude O_M is the square root of the transmission coefficient measured by absorptiometry, and whose phase O_P is the retardation measured by refractometry. In any plane z_j beyond z_1 the object changes the amplitude from F_{oj} to

$$O_j = H_1 O F_{01} F_{1j} \qquad \ldots\ldots(3)$$

The relative intensity, which indicates the effect of the object, is

$$R = IO_i / IF_{0j} \qquad \ldots\ldots(4)$$

The image plane conjugate with z_1 will be called z_3, and between z_1 and z_3 there is assumed to be a plane z_2 conjugate with z_0, containing a second aperture stop S_2. Applying equation 3 to the image plane 3 gives

$$O_3 = H_1 O F_{01} F_{13} \qquad \ldots\ldots(5)$$

where F_{13} is the diffraction pattern surrounding the image in plane 3 of a point in plane 1. For any point (r_3,ϕ_3) where O is constant over the whole of F_{13}, which usually covers a very small region, O commutes with H_1, so that O_3 becomes

$$O_3 = O H_1 F_{01} F_{13} = O F_{03} \qquad \ldots\ldots(6)$$

where O_3 and O refer to conjugate points. Under these conditions the relative intensity in equation 4 reduces to

$$R = O_M^2 \qquad \ldots\ldots(7)$$

and depends therefore solely on the intensity transmission coefficient O_M^2 of the object, not on the retardation O_P. The image of a transparent uniformly refractile object $O = e^{iO_P}$, viewed by an ordinary microscope, is therefore completely invisible, however it is illuminated, except within the region of the image covered by the diffraction pattern of its edges.

If it is desired to make the uniform part of an object visible, the diffraction pattern of the edges can be broadened by operating on the aperture of the objective. The pattern can be so arranged that it has a sharp central peak surrounded by a diffuse halo, as in the case discussed by ZERNIKE[18] where a small part in the middle of the aperture is operated on. This causes the middle of a fair-sized object to be visible, provided that the source S_B is small enough to be focused on the modified region of the aperture, without appreciable loss of resolving power. Any such focal methods in which visibility is improved by operating on the objective aperture will be called here aperfocal. The term covers the ordinary dark ground method in all its forms, and Zernike's phase-contrast method. It implies division of the aperture into two parts which give different treatments to the corresponding components of the beam of light, a process which is familiar in interferometry where it is called aperture splitting and is exemplified by the Rayleigh refractometer as mentioned earlier.

There is an alternative way of splitting a beam into components, which is called amplitude splitting, and focal methods employing it will be called here amplifocal. It is exemplified by the various interference microscopes, though the latter term is ambiguous because interference can be produced by aperture splitting too. Amplitude splitting is achieved by means of partial reflectors, which can split a beam into exact replicas whose amplitudes can have any relation subject only to conservation of energy. When two such replicas are suitably directed against opposite sides of a partial reflector they undergo a linear transformation giving two other replicas, one of which may be given zero amplitude so that the other can have the original amplitude if desired. This principle is familiar in ordinary interferometry[3], but its application to high power microscopy is more recent[19-23]. If the object is between a splitting and a recombining partial reflector, so that it affects only one component beam, the relative intensity in the image plane is transformed to

$$R_T = I(T_O O_3 + T_F F_{03})/I_5 F_{03} \qquad \ldots . (8)$$

where T_O, T_F are complex numbers independent of r, ϕ and subject only to the energy condition $T_{OM} + T_{FM} \leqslant 1$. For points in the image where equation 6 is valid *i.e.* where the object O is sufficiently uni-

form within the area of the point diffraction pattern F_{13} to be accurately reproduced in the image O_3, equation 8 reduces to

$$R_T = I(T_O O + T_F)F_{03}/IF_{03} = (T_O O + T_F)^2_M IF_{03}/IF_{03} = (T_O O + T_F)^2_M \quad \ldots.(9)$$

$$= T^2_{OM}O^2_M + T^2_{FM} + 2.T_{OM}T_{FM}O_M \cos(O_P + T_{OP} - T_{FP}) \quad \ldots.(10)$$

which reduces further to

$$R_T = T_B + T_E\{\cos T_P(\cos O_P - 1) + \sin T_P \sin O_P\} \quad \ldots.(11)$$

when the object is transparent ($O_M = 1$). Subscripts M, P refer to magnitude and phase as usual, and

$$T_B = T^2_{OM} + T^2_{FM} + T_E \cos T_P \quad \ldots.(12)$$

is the background intensity, while

$$2T_E = 4T_{OM}T_{FM} \quad \ldots.(13)$$

is the efficiency *i.e.* the fraction of object-modulated energy collected, and

$$T_P = T_{OP} - T_{FP} \quad \ldots.(14)$$

can be chosen to make the relative intensity depend on either sin O_P or cos O_P or on both. If $T_P = \pm\pi$ and $T_{OM} = T_{FM} = \frac{1}{2}$, so that

$$2T_E = 1 \quad \ldots.(15)$$

equation 11 becomes

$$R_T = (1 - \cos O_P)/2 \sim O^2_P/4 \text{ when } O_P \ll \pi/2 \quad \ldots.(16)$$

This may be called the efficient cosine case. The efficient sine case

$$T_P = (n \pm \tfrac{1}{2})\pi, 2T_E = 1 \quad \ldots.(17)$$

is similarly

$$R_T = (1 + \sin O_P)/2 \sim (1 + O_P)/2 \text{ when } O_P \ll \pi/2 \quad \ldots.(18)$$

Inefficient sine and cosine cases are obtainable with other values of T_{OM}, T_{FM}, and hybrid cases with other values of T_P. When $O_P \ll \pi/2$ the sine case is linear in O_P while the cosine case is quadratic. This was first noticed by ZERNIKE[24] who was led to it by consideration of Abbé spectra and who managed by means of his phase ring to produce a partial sine effect in an aperfocal system. The linearity in the sine case distinguishes positive from negative retardations and gives more object-modulated energy when O_P is small. This is important for photography of small objects in Brownian movement ; but for visual observation the visibility is mainly determined, within the region of validity of the Weber-Fechner law, by the visual contrast, log (image intensity/background intensity). The visual contrast is infinite in the cosine case and tends to zero for small O_P in the sine case, so that the

term phase contrast for Zernike's method is unfortunate. Actually he recommends a hybrid case corresponding to

$$T_P = \pi + O_{P_{max}} \qquad \ldots\ldots(19)$$

where $O_{P_{max}}$ is the greatest value of O_P that it is desired to observe. This gives infinite contrast for the one chosen value of O_P. The Weber-Fechner law, however, is not the whole story in questions of visibility, and further study from this point of view would probably make the sine case, or Zernike's hybrid, seem more favourable. The pure sine case cannot be produced by Zernike's method unless the phase ring is less absorbing than the rest of the aperture, as shown in Appendix IV. Zernike's method in the form that he recommends is highly inefficient in the sense that $2T_E \ll 1$ if R_T refers to full aperture illumination, because his source only fills a small fraction of the aperture. The efficiency can be calculated roughly in any specific case with the aid of Appendix IV.

AMPLIFOCAL METHODS

The theoretical simplicity of amplifocal methods, resulting from their power of achieving exactly the transformation expressed in equation 8, is offset by considerable technical difficulty in making the recombined replicas coincide exactly. For instance, although the MERTON method[21, 22] produces a series of replicas which have passed different numbers of times through the object, as is necessary for visibility in the amplifocal method, they cannot be superposed exactly unless the thickness of the object is negligible compared with the wavelength. If the object has thickness t there is a longitudinal array of replicas of the illumination system and object, spaced $2t$ apart when measured in the medium of refractive index μ surrounding the object. The spacing causes the relative retardations of rays in the recombined replicas going to the objective to depend on the angle θ that they make with the z axis. The relative intensity in the image plane, due to an elementary ring of aperture corresponding to θ is given by

$$R_T = \frac{(1-b)^2}{1 + b^2 - 2b\cos(2O_P + 2\mu t\cos\theta)} \qquad \ldots\ldots(20)$$

where b is the reflection coefficient of the partial reflectors. The main drawback of equation 20 as compared with equation 8 is that the periodic dependence on θ causes the object to have reversed effects in different parts of the aperture, so that the average effect over the whole aperture is small unless t is very small. MERTON[22] met this difficulty by inserting a zone plate, cutting out all but certain angles of illumination. The primitive low aperture form of the Merton method, con-

sisting merely in using rhodiumized slides and coverslips in an ordinary microscope, is simple and elegant, but the need of a zone plate at high apertures spoils both its efficiency and its practical convenience. The trouble is due to the use of multiple-beam interferometry, producing as it does an inseparable series of replicas in different positions.

The standard way of making replicas coincide exactly is by means of double-beam interferometry, where two replicas are isolated and handled separately, as in the Michelson interferometer[3]. This has been applied to microscopy by LINNIK[20], and developed by KINDER[19], giving an instrument in which any values of T_O and T_F in equation 8 can be obtained at will with full aperture for both illumination and viewing *i.e.* with perfect efficiency. The Linnik method should be ideal in reflectometry (metallography), though its adjustment, as with most double beam instruments, requires precision in manufacture and skill in operation. In refractometry the Linnik method has the slight drawback that when a transparent object rests on the mirror the light passes twice through it and gives two contiguous image planes, mutually reversed in the z direction, which confuse each other. The two image planes can be made coincident at any chosen level in the object, if the mirror immediately below the object is replaced by a second, inverted objective at the centre of a concave mirror whose surface coincides with the image plane of the second objective. The returning light then carries an image coincident with the object, and gives a single final image plane. This would appear to lead to an efficiency greater than unity, as measured by equation 13, because all the light scattered both forward and backward in two passages through the object is used, but the question is full of pitfalls which will not be discussed here. Ideally this modified method would require four perfect objectives, two in the object beam and two in the comparison beam. In principle the two latter could be replaced by a fairly simple low aperture lens system, at least for the centre of the field. The main trouble with such a method, which has been tried roughly[17] but deserves further study, is the complexity of the adjustments. Another possible way of avoiding the slightly defocused second image is to defocus it still more, thereby blurring it out completely *e.g.* by placing a transparent film between the object and the supporting mirror. This also deserves study. The same principle is used in a microscope[23] based on the cyclic form of Michelson interferometer. Here the two replicate beams pass in opposite directions through opposed objectives with the object in between. There are then two image planes whose distance apart can be altered at will. If they coincide they will mainly interfere and become invisible unless the object is thin compared with the wavelength. One image plane is therefore usually defocused enough to

make its disturbing effect barely perceptible. This cyclic system is fairly easy to adjust except that the objectives need a very fine and rigid centring adjustment. It is as easy to use with white light as with monochromatic, so that the continuous background of a high-pressure mercury lamp does not disturb it. Both this method and the Linnik method need high grade objectives, but these are necessary for good image formation in all methods, and a method which automatically detects a poor objective or bad alignment is sometimes positively advantageous.

When the splitting and recombining partial reflector in an amplifocal method are one and the same, as in the Michelson interferometer, one of the two emerging transformed replicas has T_{OM} and T_{FM} in equation 10 automatically equal, as required for maximum efficiency. In the cyclic type[23] this balanced replica returns towards the source and cannot be used unless another partial reflector separates part of it from the illuminating beam travelling in the opposite direction. This is inefficient, and it is better to use the other replica while balancing it by making the reflector exactly semireflecting. In the open type[20] the unbalanced replica returns to the light source and the balanced one can be used directly. Another difference between the cyclic and the open type is that in the former T_P of equation 14 is uniquely determined by the nature of the partial reflector, and can only be altered by using a different one. It is this feature that makes the cyclic type easy to use with white light and with varying object thickness, because T_P does not need continual readjustment; but a method of changing T_P continuously is desirable in some applications, and does not at present exist for the direct cyclic type. With a reversed cyclic type[23], which has been studied to some extent, interference fringes can be formed across the image as with the open type. Here all values of T_P are available simultaneously in different parts of the field. In the direct cyclic type the fringes are located in the objective, and must be broadened till the central one fills it completely to obtain full efficiency. Fringe width is easily controllable in this method by tilting the pair of objectives relatively to the mirrors, but the centring must first be adjusted till the fringes are straight, otherwise they cannot be broadened sufficiently. A theoretical prediction of T_P in the cyclic method can be made from the optical constants of the film materials. Using formulae of CROOK[25] for any film, simple or composite, it is readily shown that $-\exp(iT_P)$ is equal to the squared product of the amplitude transmission coefficients. If these are real (perfect dielectrics), the cosine case, equation 16, should be obtained. An experimental aluminium semi-reflector (imperfect conductor) gives, as it should, something approaching the sine case. ZEEMAN[26] obtained

a wide range of T_P in experiments with metals which were aged and chemically treated in various ways.

It may be found that the most satisfactory amplifocal microscope for refractometry is one where all light travels in only one direction, as in the Jamin interferometer[46]. This would give a single image, continuously variable T_P, and freedom from stray light.

It is well known that any two dimensional amplifocal method, where straight comparison fringes can be produced in the image plane with the object removed, can be made into a one dimensional self-plotting method by arranging that the comparison fringes are parallel to the gradient of retardation in the object. They will then be distorted by the object so as to plot its retardation in the perpendicular direction, as in the astigmatic form of Rayleigh refractometer mentioned above. This is useful in microscopy for measuring the retardation of moderate sized uniform objects. If the comparison fringes are transverse their spacing measures the gradient of retardation. If the latter is due to a wedge-shaped object of constant refractive index the latter quantity can be measured *e.g.* KINDER[19] who seems in this to have anticipated the Merton method when measuring the refractive index of erythrocytes.

APERFOCAL METHODS

The great practical convenience of aperfocal methods is demonstrated by their widespread use. They are mostly varieties of the Foucault-Toepler schlieren methods. Their long history is discussed by LAMM[5] who was the first to adapt the schlieren method to the ultracentrifuge in the form of his slit method[27]. Later TISELIUS *et alii*[28] used it somewhat differently, and other workers introduced the astigmatic forms already mentioned[29-33, 35-37]. The astigmatic form of Rayleigh refractometer mentioned above[16] is peculiar in being not a schlieren method but an aperfocal interference method. It is also peculiar in that the aperture splitting is parallel to the gradient of retardation, so that the effective aperture and resolving power in the direction of the latter are maximal. Also it plots the retardation itself, not the gradient. This property is shared by the astigmatic prismatic method mentioned above[17] which is neither aperfocal nor amplifocal and may be regarded as the sole representative of a third type of focal method, the prisma-focal, defined as a system where the local refractive index along one dimension is measured by means of a proportionate transverse gradient of retardation, imparted by the prismatic shape of the object.

In one dimension the term schlieren method usually includes any aperfocal method where the aperture is split by stops transversely to the variation of retardation, but ZERNIKE[4] restricts the term in microscopy to the special case where the zero order Abbé spectrum and all

those on one side of it are transmitted. According to Zernike most two dimensional aperfocal methods seem to give variously distorted measures of retardation. On the other hand the one dimensional forms are usually supposed to give a fairly accurate measure not of retardation but of its gradient. This discrepancy has nothing to do with dimensions, but depends on conditions such as steepness and extent of gradients, focal length, and shape of stop. The importance of the gradient of retardation in dark ground and phase-contrast microscopy under some conditions has been pointed out by HALLIMOND[38]. In comparing the two methods experimentally he gives some evidence for the view that changing the phase of the background (phase contrast) instead of suppressing it (stop contrast as Hallimond calls it *i.e.* ordinary dark ground with a narrow stop) does not affect the essential image pattern, but merely changes the presentation of it from the cosine to a partial sine type (equations 16, 18). A different opinion, that phase contrast somehow improves the visibility of the middle of a uniform object as compared with a dark ground system with the same shape of aperture, is sometimes encountered among users. The confusion seems to result from the fact that Zernike made the background beam annular, with intermediate radius, at the same time as he introduced phase contrast. Further, Zernike's calculations refer not to the annular case, which requires the integral given in Appendix III, but to a small central disc. The conditions under which the middle of a uniform object can be made visible when the illumination aperture is large enough to be efficient, are of practical importance, but will probably require much tedious computation before they are fully understood. The diffuse diffraction pattern with a sharp central peak, needed to combine high resolving power with visibility in the middle of the image (see above) is given by a small central disc and probably also by a narrow ring, but the illumination aperture is minute for the former and small for the latter, and can probably only be increased by the use of multiple narrow rings if the desired diffraction pattern is to be preserved. If there is no need to see the middle of a uniform object but only small sharp discontinuities, as with the customary dark ground stops, the problem is easier, and a large central disc is probably as good as anything. Alternatively, if photographic exposure time and focal depth are unimportant, the single narrow ring or small central disc are satisfactory. A quantitative theory of the effect of source and aperture shape on the visibility of extended objects by aperfocal methods does not at present seem to exist. ZERNIKE[24] has dealt with the idealized case of an axial point source and a small axial stop or phase disc, but the practical case of an annular stop or phase ring, which requires a paraxial source, does not seem to have been published, possibly be-

cause of its intractable nature. A useful integral for this purpose, which can be applied by suitable choice of limits to the case of a uniform or prismatic central disc shaped or annular object, with a paraxial point source and a disc shaped or annular stop, all apertures being finite, is given in Appendix III, in the form of a triply infinite series of Bessel coefficients. The intensity integration of this, as in equation 2, to give the effect of a disc shaped or annular incoherent source, presents no further difficulty in principle but is laborious. The result will probably be a quadruply infinite series of Bessel coefficients. The extension to objects of continuous but otherwise arbitrary pattern, which Zernike has given for the case of an axial point source and infinite aperture, hardly seems possible with a paraxial source and finite aperture, but a paraxial infinite-aperture treatment would be quite informative.

In the one dimensional schlieren method the disc shaped or annular stop is replaced by a single or double straight edge, but the focusing is still two dimensional. There are two astigmatic (self-plotting) forms. In one, the diagonal schlieren method, using a diagonal stop and a cylindrical lens, the focusing is truly astigmatic, with the object focused in the direction of the gradient of retardation and the stop in the perpendicular direction. In the other, the schlieren scanning method[29–31], the ordinary stigmatic optical system is used and the astigmatism is introduced by mechanical scanning. The earliest form of diagonal schlieren method, due to THOVERT[39], was unsuitable for the ultracentrifuge because it demanded a wide cell (LAMM[5]), but the difficulty was overcome by PHILPOT[33]. As with all schlieren methods there is a wide choice of stops. If half the viewing aperture is obscured by a stop with a single diagonal edge, and if diffraction is blurred out by a suitably wide source, the desired graph is plotted as a sharp line of constant greyness between black and white regions. If the whole aperture is obscured except for a narrow slit, the graph is plotted as a less well defined line of maximum brightness on a dark background. This form is preferred by SVENSSON[35–37] because it is less subject to systematic error when the object absorbs or scatters light. Its inverse, with only a narrow line obscuring the aperture, is also used occasionally, and probably gives the best resolving power.

The exact theory of the one dimensional stigmatic schlieren method with axial point source is given by RAYLEIGH[14], ZERNIKE[24], and LINFOOT[1]. The two last deal with an arbitrary continuous object and infinite aperture. The integration over the light source does not seem to have been carried out. An approximate theory of the slit form of diagonal method is given by SVENSSON[35–37]. Phase contrast does not seem to have been used yet in any astigmatic methods.

CONCLUSION

It is not necessary for microscopists to grasp all the mathematical complexities of refractometry, but a skim through this review may warn them not to be too dogmatic in relating the image seen to the object studied, since any method may represent an object in a variety of ways according to circumstances. Apart from some specialized afocal methods the main interest lies in focal ones, since with them an almost undistorted object pattern can sometimes be seen. The chief forms of distortion are invisibility of the middle of a uniform object and spreading of the diffraction pattern from the non-uniform parts. These are inversely related in aperfocal methods such as dark ground and phase-contrast microscopy, so that one trouble cannot be abolished without introducing the other, though with subtle design a good image can usually be obtained in practice. A related trouble occurs in those amplifocal methods where the object gives more than one image, but it can sometimes be made harmless. In all types of focal methods there is a choice between a dark and a light background. A dark background, as in dark ground microscopy, gives high contrast with a purely quadratic dependence of intensity on small retardations. A light background, as in phase-contrast microscopy, gives low contrast except for one chosen retardation, and a mixed linear-quadratic dependence. A purely linear dependence is not used in the phase-contrast method and is only obtained efficiently by amplifocal methods. For visual observation of small retardations the dark background is probably best because of the high contrast, while for photography of particles in Brownian movement a light background of suitable phase gives more energy and therefore a shorter exposure time ; the lost contrast can be mainly recovered in the development and printing, but the benefit of high energy is only fully reaped if the efficiency is maximal, which is not so with the current design of phase ring.

The effect of distortion in aperfocal methods is easily made so extreme that they measure not retardation but its gradient. This is used deliberately in many one dimensional applications, where a knowledge of the gradient is useful.

Appendix I. Coherence

The terms coherent and incoherent are often applied to beams of light in a context where, if beams 1 and 2 give amplitudes

$$x_1 = x_{1M} \exp (ix_{1P}) \text{ and } x_2 = x_{2M} \exp (ix_{2P}) \quad \ldots . (21)$$

respectively at any point, then the combined average intensity at that point is

$$x^2{}_{1M} + x^2{}_{2M} \quad \ldots . (22)$$

for incoherent light and

$$x^2_{1M} + x^2_{2M} + 2x_{1M}x_{2M} \cos (x_{1P} - x_{2P}) \quad \ldots.(23)$$

for coherent light. This conception may be generalized to intermediate cases where the combined intensity is

$$x^2_{1M} + x^2_{2M} + 2c_{12}x_{1M}x_{2M} \cos (x_{1P} - x_{2P}) \quad \ldots.(24)$$

and c_{12} is a number between 0 and 1 which may be defined as the coherence. If 1 and 2 refer to different primary sources, or to different parts of one primary source, then $c_{12} = 0$, but if they refer to secondary sources, then c_{12} depends on the distance between points 1 and 2 and on the angle subtended by the primary source. If the secondary sources are in the condensor image of a primary source, then c_{12} depends on the aperture of the condensor, because this determines the overlap of diffraction patterns of adjacent images.

The controversy between advocates of Koehler and of critical illumination turns on the question of coherence and shows the need for a satisfactory general treatment. One possible approach is to split the coherence number c_{12} into the product $c_1 \times c_2$ where c_1 and c_2 are non-numerical symbols belonging to the points 1 and 2. The algebra of these symbols is a fascinating study but is difficult to develop without contradictions. Until some such method is available, the conventional one must be used *viz* to work through the complete optical system with light from one point in the source, establishing the intensity at an image point as a function of the position of the source point, and then integrating the intensity for all source points. Usually a single source point involves so much labour that the final integration over the source is omitted. This is the cause of many of the discrepancies between theory and practice in optics.

In connection with coherence Professor Zernike has drawn my attention to two of his papers[40, 41]. The former paper has led him to discover the following elegant theorem which he has invited me to mention :

> For a point source imaged at the centre of a phase strip of any form, the intensity distribution in the halo round a small patch in the object represents the amplitude of the corresponding diffraction image, whereas a light source whose image covers the whole of the phase-strip gives rise to a halo corresponding to the intensity of the diffraction image.

Appendix II. Definition of Complex Amplitude

The scalar complex amplitude F is a complex number expressible in the form

$$F = F_S + iF_C = F_M \exp (iF_P) = F_M \cos F_P + iF_M \sin F_P \quad \ldots.(25)$$

where F_S and F_C are real numbers expressing the amplitudes of the oscillations $F_S \sin kt$ and $F_C \cos kt$ respectively, $k = 2\pi/\lambda$, λ is the wavelength, and t the time. The subscripts S, C, M, P will be used similarly for all other complex numbers that occur. A unit expanding spherical wave is defined as one for which $F = \exp(ir)/r$ at distance r from its origin. All distances are expressed with $\lambda/2\pi$ as unit, so that k can be omitted from all formulae.

The combined oscillation $(F_S \sin kt/x + F_C \cos kt)$ is expressible in the form $F_M \sin(kt + F_P)$ where

$$F_M = (F^2{}_S + F^2{}_C)^{\frac{1}{2}}$$

is its total real amplitude and

$$F_P = \tan^{-1}(F_C/F_S)$$

is its phase angle. The intensity of the combined oscillation, as detected by the eye or by a photographic plate, is

$$F^2{}_M = F^2{}_S + F^2{}_C = F\bar{F} \qquad \ldots.(26)$$

where $\bar{F} = F_S - iF_C = F_T \exp(-iF_P)$ is the complex conjugate of F. The term amplitude will be used to mean scalar complex amplitude unless qualified by adjectives such as real.

A single light wave can only be properly represented by a transverse vector complex amplitude consisting of two mutually perpendicular complex components lying in the wave front. This introduces an awkward asymmetry, since a radially symmetrical transverse vector wave is impossible. The difficulty is usually glossed over, but it is dealt with by LUNEBERG[42] who shows that when polarization is negligible, the observable effects of a light wave can be represented by a scalar complex amplitude which he defines by averaging the intensity over all directions of polarization. Further difficulties arise when the amplitude F_n on the surface n has to be calculated from F_{n-1} on the surface $n-1$ using Huygens' principle or some modern counterpart[16, 42-44], but these difficulties will not be discussed further here.

Appendix III. Image of a uniform disc with paraxial point source but circular symmetry

The light going through a uniform disc shaped object forms the same pattern as that through a stop of the same diameter with no object, but with phase differing by O_P. It can therefore be calculated from equation 1 and added, after phase change by O_P, to the background light determined from the same equation with different integration limits. In the example chosen here, equation 1, expanded to the full, is

$$F_{03} = H_2 F_{23} H_1 F_{12} F_{01} \qquad \ldots.(27)$$

where, with perfectly corrected lenses

$$F_{j,j+1} = \exp\,(-\rho_{j,j+1}\cos\phi_{j,j+1}) \qquad \ldots.(28)$$

$$\phi_{j,j+1} = \phi_{j+1} - \phi_j \qquad \ldots.(29)$$

$$\rho_{j,j+1} = r_j r_{j+1}/z_{j,j+1} \qquad \ldots.(30)$$

and $z_{j,j+1}$ is the distance between the planes in the equivalent thin-lens system. With circular stops

$$H_j = \int_0^{r_j}\int_0^{2\pi} \ldots. d\phi_j dr_j \qquad \ldots.(31)$$

Then

$$F_{02} = H_1 F_{12} F_{01}$$

$$= \int_0^{r_1}\int_0^{2\pi} \exp\left\{-r_1\left(r_2\cos\phi_{12}/z_{12} + r_0\cos\phi_{01}/z_{01}\right)\right\} d\phi, dr \qquad \ldots.(32)$$

Put

$$r_1\{r_2\cos\phi_{12}/z_{12} + r_0\cos\phi_{01}/z_{01}) = \rho_{02}\cos\,(\phi_1 - \psi_{02}) \ldots.(33)$$

where

$$\rho_{02} = (\rho_{01}^2 + \rho_{12}^2 - 2\rho_{01}\rho_{12}\cos\phi_{20})^{\frac{1}{2}} \qquad \ldots.(34)$$

$$\phi_{20} = \phi_0 - \phi_2 = \phi_{23} - \phi_{03}\ etc \qquad \ldots.(35)$$

$$\psi_{02} = \tan^{-1}[(r_2/z_{12}.\sin\phi_2 + r_0/z_{01}\sin\phi_0)/(r_2/z_{12}\cos\phi_2 + r_0/z_{01}\cos\phi_0)] \qquad \ldots.(36)$$

Combining equations 32 and 33 gives

$$F_{02} = \int_0^{r_1}\int_0^{2\pi} \exp\left\{-\rho_{02}\cos(\phi_1 - \psi_{02})\right\} d\phi_1 dr_1 = \int_0^{r_1} r_1 \mathcal{J}_0(\rho_{02})$$

$$= r_1^2 \mathcal{J}_1(\rho_{02})/\rho_{02} \qquad \ldots.(37)$$

where $\mathcal{J}_j$ is the Bessel coefficient[12] of order j. This is the usual diffraction formula for a point image, but with transformed coordinates. Inserting it in equation 27 gives

$$F_{03} = r_1^2\int_0^{r_2}\int_0^{2\pi} \exp\,(-\rho_{23}\cos\phi_{23})\mathcal{J}_1(\rho_{02})\rho_{02} d\phi_2 dr_2 \qquad \ldots.(38)$$

Equation 34, using the Neumann-Graf addition theorem[12], and expanding the denominator by the binomial theorem gives

$$\mathcal{J}_1(\rho_{02})/\rho_{02} = \frac{\sum_{m=-\infty}^{m=+\infty} \mathcal{J}_{1+m}(\rho_{01})\mathcal{J}_m(\rho_{12})\exp\,(mi\phi_{20})}{\left\{\rho_{01} - \rho_{12}\exp\,(-i\phi_{20})\right\}}$$

$$= \sum_{m=-\infty}^{m=+\infty} \mathcal{J}_{1+m}(\rho_{01})\mathcal{J}_m(\rho_{12})\ \frac{\exp\,(mi\phi_{20})}{\rho_{01}}\ \sum_{n=0}^{n=\infty}\left(\frac{\rho_{12}\exp\,(i\phi_{20})}{\rho_{01}}\right)^n \qquad \ldots.(39)$$

The above procedure is only valid provided that $\rho_{12} < \rho_{01}$ *i.e.* that the point chosen in plane 2 lies nearer the centre than the image of the point chosen in plane 0. If $\rho_{12} > \rho_{01}$ their roles in equation 39 are interchanged so that the integration limits must be split into two parts corresponding to the regions inside and outside the field stop or the object, whichever is being considered.

Substituting equations 39 and 35 in equation 38 and rearranging, gives, with $\rho_{12} < \rho_{01}$,

$$F_{03} = \frac{r_1^2}{r_{01}} \sum_{m=-\infty}^{m=+\infty} \mathcal{J}_{1+m}(\rho_{01}) \sum_{n=0}^{n=\infty} \exp\{i(n-m)\phi_{03}\} \times$$

$$\times \int_0^{r_2} r_2 \left(\frac{\rho_{12}}{\rho_{01}}\right)^n \mathcal{J}_m(\rho_{12}) \int_0^{2\pi} \exp\{i(\phi_{23} - \rho_{23}\cos\phi_{23})\} d\phi_2 dr_2$$

$$= \frac{r_1^2}{\rho_{01}} \sum_{m=-\infty}^{m=+\infty} \mathcal{J}_{1+m}(\rho_{01}) \sum_{n=0}^{n=\infty} \exp\{i(n-m)\phi_{03}\} \int_0^{r_2} r_2 \left(\frac{\rho_{12}}{\rho_{01}}\right)^n$$

$$\mathcal{J}_m(\rho_{12}) \mathcal{J}_{m-n}(\rho_{23}) dr_2 \qquad \ldots\ldots(40)$$

By repeated integration by parts in either of two directions, after multiplying one part and dividing the other by suitable powers of r_2, the integral in equation 40 can be expressed in two alternative forms. They are both convergent, but when one converges rapidly the other converges slowly. The change over occurs when $\rho_{23} = \rho_{12}$ *i.e.* when the point chosen in plane 3 is the same distance from the centre as the image of the point chosen in plane 1. The series in question are

$$\int_0^{r_2} r_2 \left(\frac{\rho_{12}}{\rho_{01}}\right)^n \mathcal{J}_m(\rho_{12}) \mathcal{J}_{m-n}(\rho_{23}) dr_2$$

$$= \frac{r_2^2}{\rho_{12}} \left(\frac{\rho_{12}}{\rho_{01}}\right)^n \sum_{p=0}^{p=\infty} \mathcal{J}_{m+p+1}(\rho_{12}) \mathcal{J}_{m-n+p}(\rho_{23}) \left(\frac{\rho_{23}}{\rho_{12}}\right)^p \qquad \ldots\ldots(41)$$

for $\rho_{23} < \rho_{12}$

$$= \frac{-r_2^2}{\rho_{23}} \left(\frac{\rho_{12}}{\rho_{01}}\right)^n \sum_{p=0}^{p=\infty} \mathcal{J}_{m-p}(\rho_{12}) \mathcal{J}_{m-n-p+1}(\rho_{23}) \left(\frac{\rho_{12}}{\rho_{23}}\right)^p \qquad \ldots\ldots(42)$$

for $\rho_{23} > \rho_{12}$

A similar procedure is possible when $\rho_{12} > \rho_{01}$, with slight modifications. Equations 39, 41 and 42 can be combined in the appropriate way according to the integration limits, to give F_{03} for any chosen case. Separate limits are chosen for the object disc and the image stop and the results are combined after introducing the object retardation O_P, assumed constant. A prismatic object disc is equivalent to a change in $r_0\phi$, and causes no difficulty till the final intensity-integration I is attempted, when it introduces asymmetry.

Appendix IV. Efficiency and Linearity of Aperfocal Methods

Pending the development of the complete theory discussed above a qualitative general theory of efficiency and linearity, as defined by equations 10–18 for amplifocal methods, can be given for aperfocal methods such as the phase-contrast method, with the dark ground method regarded as the limiting case where the phase film transmission is zero. The term phase film is preferred to phase ring for this purpose, since it implies no particular shape. Phase plate is unsuitable because it includes the supporting structure.

Returning to equation 1 the parts S_{B2} and S_{D2} of the aperture S_2 are now phase films with complex amplitude transmission coefficients $\mathcal{Z}_B$ and $\mathcal{Z}_D$ respectively, each being constant over its own part of the aperture. The practical case is assumed where the stop S_0 defining the source has S_{B0} conjugate with S_{D2}, so that in the absence of an object, and neglecting diffraction, the whole of S_{D2} and none of S_{B2} is illuminated. The effect of these assumptions is to split O_3 of equation 5 into two parts, O_{B3} and O_{D3}, where

$$\begin{aligned} O_{B3} &= \mathcal{Z}_B H_{B2} F_{23} H_1 F_{12} O F_{01} \\ O_{D3} &= \mathcal{Z}_D H_{D2} F_{23} H_1 F_{12} O F_{01} \end{aligned} \qquad \ldots\ldots(43)$$

with H_{B2}, H_{D2} representing Huygens' integrations over the partial apertures S_{B2}, S_{D2}. The total image amplitude is therefore

$$O_3 = O_{B3} + O_{D3} = (\mathcal{Z}_B H_{B2} + \mathcal{Z}_D H_{D2}) F_{23} H_1 F_{12} O F_{01} \qquad \ldots\ldots(44)$$

and the image intensity is

$$I_B O_3 = I_B (\mathcal{Z}_B H_{B2} + \mathcal{Z}_D H_{D2}) F_{23} H_1 F \qquad \ldots\ldots(45)$$

where I_B is the intensity operator of equation 2 acting only over the region S_{B0}.

If the system has circular symmetry except for the variable position (r_0, ϕ_0) of a point on the source, the quantities O_{B3}, O_{D3} of equations 43 and 44 are special cases of the integral given in *Appendix III*. An exact result can therefore be obtained in this special case, given time and courage, since the further integration I_B and the computation are not more difficult than the rest but only laborious. The qualitative result can, however, be seen without this labour, even in asymmetric cases. The fact that S_{B0} and S_{D2} are conjugate implies that

$$F_{B03} = \mathcal{Z}_B H_{B2} F_{23} H_1 F_{12} F_{01} \sim 0 \qquad \ldots\ldots(46)$$

$$F_{D03} = \mathcal{Z}_D H_{02} F_{23} H_1 F_{12} F_{01} \sim \mathcal{Z}_D F_{03} \qquad \ldots\ldots(47)$$

This property of the integrals is shown subtly in equation 44 by a change over between two alternative expansions when r_2 changes from the region S_{B2} to S_{D2}. Subtracting the sum of equations 46 and 47

from equation 44 and rearranging gives

$$O_3 = \mathcal{Z}_D F_{03} + (\mathcal{Z}_B H_{B2} + \mathcal{Z}_D H_{D2}) F_{23} H_1 F_{12} (O - 1) F_{01} \quad \ldots . (48)$$

If now O is constant over the whole non-vanishing part of the modified point diffraction pattern $(\mathcal{Z}_B H_{B2} + \mathcal{Z}_D H_{D2}) F_{23} F_{12}$ for some image point (r_3, ϕ_3), the object is invisible at that point as in equation 7, but the modified diffraction pattern is more extensive than the unmodified one F_{13}, so that a uniform object is visible further from its edges.

To find the linearity and efficiency corresponding to equations 10–18 of the amplifocal case we must consider a point object, since there is no simple expression for a non-uniform extended object in the aperfocal case. For a point object, $(O-1)$ in equation 48 is zero except at the point (r_1, ϕ_1) where the object is located. Hence equation 48 becomes

$$O_3 = \mathcal{Z}_D F_{03} + (\mathcal{Z}_B H_{B2} + \mathcal{Z}_D H_{D2}) F_{23} F_{12} (O-1) F_{01} r_1 d\phi_1 dr_1 \quad \ldots . (49)$$

The surface element $dA_1 = r_1 d\phi_1 dr_1$ measures the infinitesimal extent of the point object and determines how much object-modulated light is available. Since $(O - 1)dA_1$ commutes with H_2, equation 49 becomes

$$O_3 = \mathcal{Z}_D F_{03} + (O-1) dA_1 . (\mathcal{Z}_B H_{B2} + \mathcal{Z}_D H_{D2}) F_{23} F_{12} F_{01} \quad \ldots . (50)$$

It will now be assumed as a qualitative approximation that the peak amplitude in the diffraction pattern at its mid-point (r_3, ϕ_3), is

$$(O - 1)(\mathcal{Z}_B A + \mathcal{Z}_D (1 - A)) K F_{03} \quad \ldots . (51)$$

where K is a constant depending mainly on dA_1, and A, $(1 - A)$ are the areas of S_{B2}, S_{D2} expressed as fractions of the total. Substitution in equation 50 gives

$$O_3 = [O(\mathcal{Z}_B A + \mathcal{Z}_D (1 - A)) - (\mathcal{Z}_B A - \mathcal{Z}_D A)] F_{03} \quad \ldots . (52)$$

Comparing equation 52 with equation 9 for the amplifocal case gives

$$\begin{aligned} T_O &= \mathcal{Z}_B A + \mathcal{Z}_D (1 - A) \\ T_F &= - A(\mathcal{Z}_B - \mathcal{Z}_D) \end{aligned} \quad \ldots . (53)$$

but R_T in equation 9 must be multiplied by the fraction A when calculating the efficiency because only that fraction of the source is used. In the usual phase-contrast case $\mathcal{Z}_B = 1$, $\mathcal{Z}_D = i\mathcal{Z}_{DM}$, and $A \sim 1$ because the phase ring is narrow. Then

$$\begin{aligned} T_O &= 1 \\ T_F &= -1 + i\mathcal{Z}_{DM} \end{aligned} \quad \ldots . (54)$$

so that $T_P = \pi - \tan^{-1} \mathcal{Z}_{DM}$

Equation 54 gives the cosine case, as equation 16 when $\mathcal{Z}_{DM} = 0$ (dark ground illumination), but making $\mathcal{Z}_{DM} = 1$ (non-absorbing phase ring) makes

$$T_P = 3\pi/4 \quad \ldots . (55)$$

which is only halfway towards the pure sine case of equation 18. The latter can only be achieved in aperfocal methods by making $Z_{BM} < Z_{DM}$ *i.e.* by the inefficient procedure of absorbing some of the object beam instead of the background.

REFERENCES

[1] LINFOOT, E. H. *Proc. roy. Soc.A* 186 (1946) 72
[2] TWYMAN, F. *Phil. Mag.* 35 (1918) 49
[3] WOOD, R. W *Physical Optics* New York, 1934
[4] ZERNIKE, F. *Physica Eindhoven* 9 (1942) 686
[5] LAMM, O. *Nova Acta R. Soc. Sci. Uppsala*, Ser. IV 10 (1937) No. 6
[6] SVENSSON, H. *Nature, Lond.* 161 (1948) 234
[7] GOUY, A. *C.R. Acad. Sci., Paris* 90 (1880) 307
[8] COULSON, C. A., COX, J. T., OGSTON, A. G. and PHILPOT, J. St. L. *Proc. roy. Soc.A.* 192 (1948) 382
[9] KEGELES, G. and GOSTING, L. J. *J. Amer. chem. Soc.* 69 (1947) 2516
[10] LONGSWORTH, L. G. *Ann. N.Y. Acad. Sci.* 46 (1945) 211
[11] — *J. Amer. chem. Soc.* 69 (1947) 2510
[12] WATSON, G. N. *Theory of Bessel Functions* Cambridge, 1944
[13] OGSTON, A. G. *Proc. roy Soc.A* 196 (1948) 272
[14] RAYLEIGH, LORD *Phil. Mag.* 33 (1917) 161
[15] CLAESSON, S. *Ann. N.Y. Acad. Sci.* 49 (1948) 183
[16] PHILPOT, J. St. L. and COOK, G. H. *Research* 1 (1948) 234
[17] PHILPOT, J. St. L. *Unpublished*
[18] ZERNIKE, F. *Physica Eindhoven* 9 (1942) 974
[19] KINDER, W. *Zeiss' Nachr.* (1937) *pp* 91, 101
[20] LINNIK, W. *C.R. Acad. Sci. U.S.S.R.* (1933), *p* 21
[21] MERTON, T. *Proc. roy Soc.A* 189 (1947) 309
[22] — *ibid* 191 (1947) 1
[23] PHILPOT, J. St. L. *Patent applied for 1949*
[24] ZERNIKE, F. *Physica Eindhoven* 1 (1934) 689
[25] CROOK, A. W. *J. opt. Soc. Amer.* 38 (1948) 954
[26] ZEEMAN, P. *Z. phys. Chem., Cohen-Festband* (1927) 237
[27] LAMM, O. *Nature, Lond.* 132 (1933) 820
[28] TISELIUS, A., PEDERSEN, K. O. and ERIKSSON-QUENSEL, I. B. *Nature, Lond.* 139 (1937) 546
[29] LONGSWORTH, L. G. *J. Amer. chem. Soc.* 61 (1939) 529
[30] — *Chem. Rev.* 30 (1942) 323
[31] — *Ind. Eng. Chem. Anal. Ed.* 18 (1946) 219
[32] — and MACINNES, D. A. *J. Amer. chem. Soc.* 62 (1940) 705
[33] PHILPOT, J. St. L. *Nature, Lond.* 141 (1938) 283
[34] SCHELKUNOFF, S. A. *Bell System Techn. J.* 15 (1936) 92
[35] SVENSSON, H. *Kolloidzschr.* 87 (1939) 181
[36] — *ibid* 90 (1940) 141
[37] — *Ark. Kemi Min. Geol.* 22 a (1946) No. 10
[38] HALLIMOND, A. F. *Nature, Lond.* 159 (1947) 851
[39] THOVERT, J. *Ann. Phys.* 2 (1914) 369
[40] ZERNIKE, F. *Physica Eindhoven* 5 (1938) 783
[41] — *Ned. T. Natuurk* 9 (1942) 357
[42] LUNEBERG, R. K. *Mathematical Theory of Optics* Rhode Island, 1944
[43] BAKER, B. B. and COPSON, E. T. *Huygens' Principle* Oxford, 1939
[44] COPSON, E. T. *Proc. roy. Soc.A* 186 (1946) 100
[45] AMBROSE, E. J. *J. Sci. Instrum.* 25 (1948) 134
[46] EDSER, E. *Light for Students* New York, 1921
[47] OCKENDON, F. E. J. *J. Quekett micr. Cl.* 2 Ser. IV (1946) 103

7

USE OF SOFT X-RAYS IN THE ASSAY OF BIOLOGICAL MATERIAL

A. Engström

In the study of the quantitative chemical composition of an individual cell or a group of cells in a biological tissue, great difficulties present themselves from the very beginning. These are conditioned mainly by the small size of the object to be analysed. The amounts to be analysed are within the ultramicro scale *i.e.* less than 1/1,000 μg (10^{-9} gm) so that the sensitivity of the ordinary microchemical methods of analysis is insufficient. It is therefore necessary to try to find other methods of investigation, chiefly physical, which can render possible such a quantitative cytochemical analysis.

In cytochemical analysis it is necessary to be able to analyse cytologically localized objects as small as 1 μ in diameter but for histochemical analysis larger areas may be used. One μ^3 of an ordinary biological tissue has the approximate weight 10^{-6} μg. Of that weight about three quarters is water and the lower limit for a method for determining elements and substances therefore lies at 10^{-8} to 10^{-9} μg. When determining nitrogen *e.g.* in 25 μ^3 of tissue containing about 3 per cent nitrogen the amount to be determined is of the order of magnitude of 10^{-6} μg.

To carry out chemical analyses on a cytochemical and histochemical scale it is, in principle, possible to employ two different procedures. One is to analyse such a small part of the tissue *e.g.* an intact microscopic section, that the result of the analysis can be directly correlated with the cytological picture. The other procedure is to make a preparation of a certain cell or group of cells by microdissection, which can subsequently be subjected to chemical analysis. With the use of the modern aids and appliances *e.g.* the micromanipulator, ultraviolet and phase-contrast microscopes, no great difficulties present themselves in the way of making a microdissection so that a small piece of tissue is obtained. In the latter procedure a preparation of the piece of tissue can be dissolved in a microdrop of a solvent, dried or ashed, or otherwise treated, in order to render possible a subsequent chemical analysis.

In order to be able to calculate the chemical composition per unit weight of the cell structure being analysed it is also necessary to have a method which permits the determination of the mass of small structures.

A method for mass determination is also very important because the specific gravity of different structures is little known. The amounts of tissue to be weighed are of the same magnitude as indicated above.

The absorption and emission of x-rays follow simple and clear laws. Therefore a method employing x-rays for cytochemical analysis would have many promising inherent possibilities. The great problem, however, is to analyse areas of the sample so small as to permit the solving of cytochemical problems. Special procedures have to be used as the x-rays are not affected by any type of magnifying system similar to those used for focusing radiation of longer wavelengths and charged particles in motion.

In recent years the use of soft x-rays has proved to be a valuable tool for quantitative cytochemical analysis. By absorption measurements with soft x-rays both the mass and amount of certain elements can be determined in very small structures. This review will be concerned mainly with the techniques for such chemical analyses. There will be short sections on the absorption laws and the principles for producing images by x-rays, then the technical procedures for determination of mass and elementary composition will be described and certain applications mentioned. Finally a short review of the methods of x-ray image microscopy will be included.

ABSORPTION AND SCATTERING OF X-RAYS

When a monochromatic x-ray beam of intensity I passes through a substance by which it is absorbed the intensity i of the transmitted radiation is calculated from the exponential equation :

$$i = I \exp\left(-\frac{\mu \cdot d \cdot \rho}{\rho}\right) \qquad \ldots\ldots(1)$$

In the equation, d is the thickness of the absorbing screen in cm, ρ is the specific gravity of the absorbing substance and μ the linear absorption coefficient. The product $d.\rho$ is thus the mass (q) of the sample in gm/sq cm. The quotient μ/ρ is the mass absorption coefficient and its magnitude for one and the same substance is independent of the physical state of the latter and is therefore only a function of the mass of substance per unit surface. The relationship between atomic and linear absorption coefficients is

$$\mu_a = \mu A/\rho \mathcal{N}$$

where μ_a is the atomic absorption coefficient, A the atomic weight and $\mathcal{N}$ Avogadro's number.

If several elements are present at the same time in an absorbing composite screen the following formula applies

$$i = I \exp\{-\Sigma(\mu/\rho)_k \cdot q_k\} \qquad \ldots\ldots(2)$$

Equations 1 and 2 are valid only if the incident bundle of x-rays is not too wide or divergent.

The accuracy with which the quotient between the incident and transmitted radiation has to be determined when estimating a substance by x-ray absorption measurements can be calculated. If an error of ϵ per cent is allowed in the determination of a substance X with the mass absorption coefficient μ/ρ the following equations are valid

$$\left.\begin{aligned} i &= I \exp(-\mu/\rho . x) \\ i_E &= I_E \exp\{-(\mu/\rho . x \pm \epsilon/100 . \mu/\rho . x)\} \end{aligned}\right\} \quad \ldots . (3)$$

Combining the equations we get

$$i . I_E / i_E . I = \exp(\pm \mu/\rho . \epsilon/100 . x) \quad \ldots . (4)$$

From equation 4 it can be calculated that the determination of a substance with the mass absorption coefficient of 5,000 and occurring in an amount of 10^{-4} gm/sq cm requires that the quotient between the incident and transmitted radiation must be determined with an accuracy of at least 2 per cent if the error in X shall not exceed 5 per cent.

The total mass absorption coefficient is composed of two terms, the mass transformation coefficient (also called true, fluorescent or photoelectric absorption coefficient) and the mass scattering coefficient

$$\mu/\rho = \tau/\rho + \sigma/\rho \quad \ldots . (5)$$

The value of the mass transformation coefficient τ/ρ increases rapidly with the wavelength of the radiation and atomic number of the absorbing screen. The value of the mass scattering coefficient σ/ρ is numerically small. The mass scattering coefficient includes both modified and unmodified scattering. According to the classical theory the scattering is independent of wavelength and proportional to the number of electrons per gram of the scattering substance. This did not prove to conform with experimental data.

By a modification of the classical theory for scattering KLEIN and NISHINA[1] deduced an expression which for short wavelengths did conform to the values found experimentally (COMPTON and ALLISON[2]). Limitations of techniques have restricted testing the formula to scattering at short wavelengths. VICTOREEN[3, 4] assumed that the Klein-Nishina formula also applied to longer wavelengths and empirical data seem to confirm his calculations.

For long wavelengths (soft x-rays) and elements such as occur in biological tissue the scattering can be neglected in comparison with the numerically greater photoelectric absorption (see also the section on quantitative elementary analysis).

Several formulae have been set up for the calculation of the mass

absorption coefficients. They generally have the form

$$\mu/\rho = C\lambda^n Z^m + b \qquad \ldots.(6)$$

For details see, for instance, COMPTON and ALLISON[2]. In formula 6, λ is the wavelength of the radiation, Z the atomic number of the absorber and b is an expression of the scattering. A universal absorption formula which covers all elements in a wide wavelength region was set up by JÖNSSON[5] in 1928. This general formula with tables and diagrams is included in monographs by SIEGBAHN[6] and COMPTON and ALLISON[2]. Jönsson deduced a 'true electronic absorption coefficient' τ_e which is a function of $Z\lambda$, and is independent of the nature of the absorbing screen. The relationship between the electronic and atomic photoelectric coefficients is

$$\tau_e = \tau_a/Z = \tau A/\rho N Z$$

From published tables the numerical values of the mass absorption coefficients can be calculated for all values of $Z\lambda$ between 8 and 790. Jönsson's formula can be written in the form

$$\tau_a \sim Z^{p+1}.\lambda^p \qquad \ldots.(7)$$

where τ_a denotes the atomic absorption coefficient. The exponent p is a function of both Z and λ. When the product $Z\lambda$ varies between 8 and 790 p varies continuously from 3 to 2·3.

ABSORPTION OF X-RAYS OF DIFFERENT WAVELENGTHS IN BIOLOGICAL MATERIAL

In order to perform a cytochemical analysis it is necessary to use a layer of biological material that has the thickness of a single cell or less. For the following calculations the cell layer is considered to have a thickness of 5 to 10 μ. When using x-rays the first question which arises is: in which regions of the electromagnetic spectrum is the absorption such that there is a good possibility of estimating experimentally the absorption in a screen with a mass per unit area corresponding to that of a single cell layer?

Using the formulae and tables which are referred to above, it is possible to calculate the transmission of a screen composed of a single cell layer for x-rays of a wide wavelength range. To simplify the problem the screen is assumed to contain only carbon, oxygen and nitrogen in amounts per unit area that correspond approximately to a single layer of cells. A thin collodion foil to which a small amount of an organic stain has been added serves as a model substance. The schematic representation of such a calculation is shown in *Figure 1*. The transmission in per cent is estimated for radiation of wavelengths from 0,1 to 10,000 Å (1 Å = 10^{-8} cm). From the figure it can be seen that the transmission

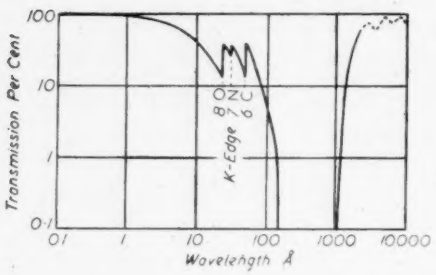


Figure 1. Transmission of electromagnetic radiation of different wavelength in a foil of coloured collodion, the mass per unit area approximating that of a single cell layer

for x-rays of short wavelengths is almost 100 per cent. In the region 0·1 to about 2 Å there is no contrast in the specimen. The region 2 to 50 Å offers good possibilities for performing absorption measurements. The wavelength region 50 to 150 Å is certainly possible to use from the point of view of transmission but the generation and measurement of the radiation is difficult in this region. When the radiation has the wavelength 150 to about 2,000 Å practically all radiation is absorbed in the specimen. Extremely high contrast of thin samples is obtained in this region. Above 2,000 Å there are of course the ultraviolet, visible and infra-red regions where the transmission is such that there are good possibilities to perform absorption measurements, but here scattering and reflection of the radiation in the sample itself is a complicating factor.

From the diagram it is obvious that for x-rays the region 2 to 50 Å is the most suitable. Radiation of these wavelengths correspond to the voltage region 6,000 to 250 volts. In this soft and extremely soft x-ray region the absorption of the radiation in air is strong and must be minimized by investigating the specimens in vacuum. A rough vacuum of about 0.01 mm Hg can be used for radiation up to 10 Å but for the longer wavelengths a high vacuum is necessary.

When performing a quantitative analysis of elements by means of x-ray absorption spectroscopy the characteristic absorption edges for the elements are used (see page 184). The biological material mainly consists of elements with low atomic numbers. The *K*-absorption edges for these elements (*Table II* p 184) fall in the region 2 to 50 Å which apparently is very favourable. In *Figure 1* the absorption edges for carbon, oxygen and nitrogen are indicated. For some elements of biological interest but of high atomic numbers the *L* and *M* edges fall within the favourable wavelength region indicated above.

Summarizing the discussion in this section it may be said that in the x-ray region from 2 to ~ 50 Å there are good possibilities for performing absorption measurements involving the transmission of a specimen consisting of a single cell layer. The characteristic absorption edges for the elements of biological interest lie within this region. Thus there are possibilities for quantitative cytochemical determination of elements by x-ray absorption spectrography.

In order to be able to perform a cytochemical analysis it is, however, necessary to measure the absorption of x-rays within a very small area of the sample. The possibilities and techniques for such measurements will be discussed in the next section.

X-RAY IMAGE FORMATION

There are two methods which can be used for determination of the absorption of x-rays within very small areas. One is to collimate x-rays to a very fine bundle and measure the absorption of this fine beam in the sample. This method has the disadvantage that it is difficult to localize accurately the structure being measured. The beam must have a diameter between 1 and 10 μ to be useful for biological investigation. The other method is to arrange an enlarged image of the sample and perform intensity measurements in this enlarged image so that the absorption can be calculated for very small areas.

As x-rays cannot be refracted in any kind of optical system analogous to that in the light or electron microscopes, enlarged images with x-rays cannot be obtained with optical, magnetic or electrical lens systems. Recently, using grazing incidence of x-rays on curved mirrors KIRKPATRICK and BAEZ[7, 8] succeeded in constructing a real x-ray microscope with a resolving power of a few microns. By the use of cylindrical or spherical crystal gratings it is possible to design a low power magnification system for x-rays. The simplest way, however, is to make a sharp shadow image which allows a considerable optical magnification. The procedure is called microradiography and the small x-ray picture is called a microradiogram.

The technique for microradiography is not new. A few years after RÖNTGEN's discovery of the x-rays, HAYCOCK and NEVILLE[9] used a radiographic procedure in an experiment which was the direct forerunner of the present day microradiographic technique. They used the radiographic technique to distinguish between gold and sodium in an alloy. The investigation was performed at unit magnification. Unfortunately the photographic emulsions available at that time were not suitable for much enlargement.

GOBY[10] published in 1913 the results of his investigations employing soft x-rays to obtain images of algae *etc* on relatively fine grained film emulsions.

It was however with the introduction of very fine grained photographic emulsions, capable of high magnification that advance of the microradiographic technique was resumed in France.

In 1930 DAUVILLIER[11] published beautiful pictures of microtome sections from tissues, taken with soft x-rays on fine grained emulsions.

The technique was mostly used for the study of alloys but the numerous papers published by LAMARQUE[12-16] and TURCHINI[17, 18] since 1936 deal with microradiography of biological material. The microradiographic technique was, however, mainly used for solving problems in metallurgy (CLARK[19-21], MADDIGAN[22] and SPROULL[23]). SIEVERT[24] and CLARK pointed out independently the possibility of adjusting the wavelength of the x-rays used with reference to the absorption edges of the elements in the sample so that the contrasts between the different elements in the sample should increase and the images allow a chemical interpretation. No quantitative measurements were performed on the microradiographic images. Microradiograms of tissues into which contrast medium had been injected were published by BOHATYRTSCHUK[25].

Beautiful pictures of leaves from plants were published by BARCLAY and LEATHERDALE[26]. BARCLAY[27, 28] has also published papers on the localization of contrast medium and iron in tissues. For a complete review of earlier papers on microradiography the reader is referred to ENGSTRÖM[29] and CROWTHER[30].

Techniques for microradiography— In *Figure 2* the different techniques for microradiography are shown. The simplest and most common experimental arrangement is shown in *Figure 2*a. The soft x-rays,

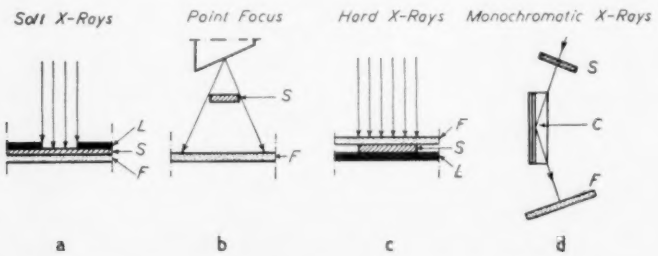


Figure 2. Different techniques for obtaining X-ray images of a specimen: S sample, F photographic film, L lead plate, C curved crystal

emitted by a tube driven at low voltage, pass through the sample *S* which lies in good contact with the fine grained photographic film *F*. A metal diaphragm *L* limits the area of the sample investigated. After exposure and development, a natural size image of the sample is obtained on the fine grained film. The microradiogram is then magnified with an optical system *e.g.* a microscope. The resolving power of the procedure depends primarily on the granularity of the film and secondarily on the inherent geometrical arrangements of the x-ray source, sample and photographic film. The errors in image formation arising from the geometrical arrangements and the resolving power for different film emulsions will be discussed later. By using an

extremely small x-ray source it is possible to obtain an enlarged shadow image of a specimen. The principle is shown in *Figure 2*b. From the point source a divergent bundle of x-rays is emitted. The sample S and the photographic film F are placed at a certain distance from each other. SIEVERT[24] obtained a small x-ray source by placing a fine pinhole in front of the x-ray tube. In order to avoid an image of the focal spot of the x-ray tube, the tube was moved during the exposure. Using an electronic lens system to focus the electrons on the anode of the x-ray tube VON ARDENNE[31] succeeded in getting a very fine focus. The voltage proposed for his x-ray shadow microscope was relatively high. However great difficulties present themselves when focusing the low energy electrons necessary for the generation of extremely soft x-rays. The total output of x-rays in such a system is very low as the load of the small area on the anode must be small. If a fine grained film emulsion such as the extremely insensitive LIPPMAN emulsion is used for recording the image the exposure times will be enormously long. The magnification depends on the mutual distances between sample, film and x-ray source. In order to obtain a primary magnification of 10 times, and if the structures to be resolved are 0·5 μ in diameter, the size of the focus should be about 0·05 μ.

*Figure 2*c shows the principle for radiography with hard x-rays. The technique was developed by TRILLAT[32] and is based on the fact that x-rays of very short wavelengths, corresponding to a voltage of several hundred kilovolts, affect the photographic emulsion to a very small extent. The secondary electrons, however, do blacken the photographic emulsion. In *Figure 2*c the hard x-rays first pass through the film F and then hit the sample S. If the sample consists of a highly absorbing substance the electrons emitted by the surface of the sample blacken the photographic emulsion. Trillat has given the name electronic radiography by reflection to that procedure. When using a thin section of organic material a lead plate L is placed behind the sample in order to increase the yield of electrons. The latter modification, applicable to biological material, is electronic micrography by transmission. The two procedures have been discussed and criticized by SEEMAN[33].

Finally in *Figure 2*d is indicated the production of a true x-ray image of a sample by the use of monochromatic radiation and a cylindrical or spherical crystal grating C. The x-rays pass through the sample S and are reflected in the crystal C and produce an image on the photographic film F. This procedure will be discussed later.

Lack of sharpness in the microradiographic image— In *Figure 3* the arrangement for microradiography is shown schematically. The maximal geometrical blurring U for a point at the distance a from the film is

$$U = fa/b \qquad \ldots.(8)$$

In the formula f is the diameter of the focal spot and b is the distance between focus and the sample. From the formula it can be seen that the blurring is small when the focal spot is small. The distance between sample and photographic film must also be small in order to keep down the blurring of the image. The ideal is to use parallel x-rays, thus a long distance between focus and sample-film. The fine grained photographic emulsions used have a very low sensitivity and therefore it is necessary to have an enormous incident x-ray intensity. Therefore the distance between focus and sample-film has to be small. The focus cannot be too small as then the total output of x-rays will be too small. The most important factor then is to keep the distance between the sample and the photographic film as small as possible when performing microradiography in the scale 1 : 1. Extremely good contact must be secured between the thin sample and the surface of the photographic film.

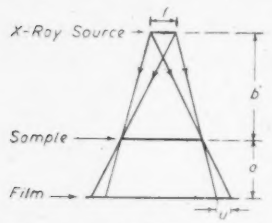


Figure 3. Lack of sharpness in the microradiographic image

Resolving power of the photographic emulsions used for microradiography— The limiting factor for the resolution in microradiography lies in the properties of photographic emulsions. In *Table I* the resolving power for some photographic emulsions is shown. The resolving power is given in lines per mm. The best emulsion is the Lippman emulsion which has a resolving power of more than 1,000 lines per mm. A microradiogram taken on Lippman film can be optically magnified about 500 times without disturbance from the grains. Investigations in the electron microscope show that the grain size in a Lippman emulsion is about 100–500 Å. Different emulsion numbers, however, show different graininess and homogenity.

Table I. Resolving Power of some Film Emulsions used for Radiography and Microradiography

Film emulsion	*Resolving power lines/mm*
Kodak X-ray film	~ 10
Agfa Printon film	~ 50
Ilford High Resolution	~ 100
Eastman Kodak Spectroscopic plate 548–0	500
Eastman Kodak Spectroscopic plate 649	≧ 1,000
Lippman film	> 1,000

When measuring the density in small areas of a film it is necessary to measure a certain number of grains depending on the statistical nature of the photographic process. Therefore the resolving power of the photographic emulsion when used for quantitative purposes is less than indicated above and is dependent on the density in the developed emulsion. The

developing conditions must be controlled. As the small x-ray images will be highly optically magnified by photomicrography it is necessary that the small films should be washed in clean water and dried in dust free atmosphere.

DETERMINATION OF THE MASS OF CELL STRUCTURES

As already mentioned in the introduction, a very important problem in cytochemical analysis is the determination of the mass of the structures under investigation. A technique for doing that would make it possible to refer other analytical findings directly to mass. Another possibility would be to refer the analytical data to a volume unit but we do not know exactly either the variation in specific density of cell structures or the effect of shrinkage introduced by cytochemical techniques. Therefore, the only correct way is to express the amount of different substances in per cent of mass as used in ordinary biochemical analyses.

First the principle for mass determination by x-ray absorption spectrography will be described. With the holder shown in *Figure 4* a microradiogram is taken of the biological specimen and a reference system simultaneously. Long wavelength continuous x-rays are used. As will be discussed later, the voltage and filtration of the radiation must be properly selected. By using a broad x-ray band, under certain experimental conditions one obtains an integration over the mass of all elements in the sample provided that the sample contains mainly organic material. If the absorption in different parts of the sample is compared with that in the reference system it is possible to calculate the mass of the respective parts. The following discussion is a brief review of a paper on a method for the determination of the mass of extremely small objects by ENGSTRÖM and LINDSTRÖM[34]. A short communication on the technique has already been published[35].

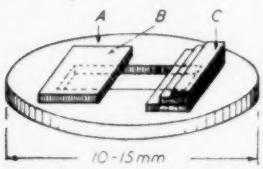


Figure 4. Technique for mounting the specimen for absorption analysis: A preparation holder, B smear or section of the biological specimen, C reference system made of thin collodion foils

Theory— The main component of ordinary biological material is protein. The percentages of carbon, nitrogen and oxygen in protein are relatively constant, and for the following calculations biological material is at first assumed to be built up of carbon, nitrogen and oxygen in these proportions. The mass absorption coefficient for this mixture is calculated. As there is a constant proportionality between the mass absorption coefficients for carbon, nitrogen and oxygen at different wavelengths on the short wavelength side of the absorption edges, it is only necessary to calculate the mass absorption coefficient

for the mixture at one wavelength. The mass absorption coefficient is then calculated for some selected wavelengths. To these basic elements small amounts of different elements, including carbon, nitrogen and oxygen are added. It is mainly the elements of biological interest, thus elements falling in the range atomic number 1 to atomic number 30, which are added. The mass absorption coefficient of the mixture is assumed to be unchanged, and therefore an approximate value of the total mass is obtained. From equation 2 the following expression is obtained

$$\ln(I/i) = \left(\frac{\mu}{\rho}\right)_{\mathrm{CNO_p}} . m_{\mathrm{CNO_p}} + \left(\frac{\mu}{\rho}\right)_{\mathrm{x}} . m_{\mathrm{x}} = \left(\frac{\mu}{\rho}\right)_{\mathrm{CNO_p}} . m_{\alpha} \quad \ldots . (9)$$

In the formula the symbols have the following meaning :

I = intensity of the incident x-rays

i = intensity of the transmitted x-rays

$\left(\frac{\mu}{\rho}\right)_{\mathrm{CNO_p}}$ = the mass absorption coefficient for C,N and O in protein for a certain wavelength

$\left(\frac{\mu}{\rho}\right)_{\mathrm{x}}$ = the mass absorption coefficient for the added element X at the same wavelength

$m_{\mathrm{CNO_p}}$ = the mass in gm/sq cm of the elements C, N and O in protein

m_{x} = the mass of the added element in gm/sq cm

m_{α} = the approximate mass of the mixture in gm/sq cm

If a systematic error ϵ per cent is allowed in the mass determination m_{α} has the value

$$m_{\alpha} = \left(1 \pm \frac{\epsilon}{100}\right) . (m_{\mathrm{CNO_p}} + m_{x}) \quad \ldots . (10)$$

where $\left(\frac{\mu}{\rho}\right)_{\mathrm{x}} > \left(\frac{\mu}{\rho}\right)_{\mathrm{CNO_p}}$ and $\left(\frac{\mu}{\rho}\right)_{\mathrm{x}} < \left(\frac{\mu}{\rho}\right)_{\mathrm{CNO_p}}$

correspond to the + and − forms respectively.

When m_{α} is eliminated from equations 9 and 10 we get

$$\frac{m_{\mathrm{x}}}{m_{\mathrm{CNO_p}}} = \frac{\pm\epsilon}{100\left\{\frac{(\mu/\rho)_{\mathrm{x}}}{(\mu/\rho)_{\mathrm{CNO_p}}}\right\} - 100 \mp \epsilon} \quad \ldots . (11)$$

When $\frac{(\mu/\rho)_{\mathrm{x}}}{(\mu/\rho)_{\mathrm{CNO_p}}} = 1 \pm \frac{\epsilon}{100}$ we obtain $\frac{m_{\mathrm{x}}}{m_{\mathrm{CNO_p}}} \rightarrow \infty$

For $1 - \frac{\epsilon}{100} < \frac{(\mu/\rho)_{\mathrm{x}}}{(\mu/\rho)_{\mathrm{CNO_p}}} < 1 + \frac{\epsilon}{100}$

the equation is not valid because between these limits m_α cannot differ ϵ per cent from the correct value. When ϵ is fixed to 5 per cent and the element X is varied from atomic number 1 to 29 the values of $\frac{m_x}{m_{CNOp}}$ in equation 11 are graphically represented in *Figure 5*. Such graphs are calculated for several wavelengths but in *Figure 5* only calculations for two wavelengths are shown.

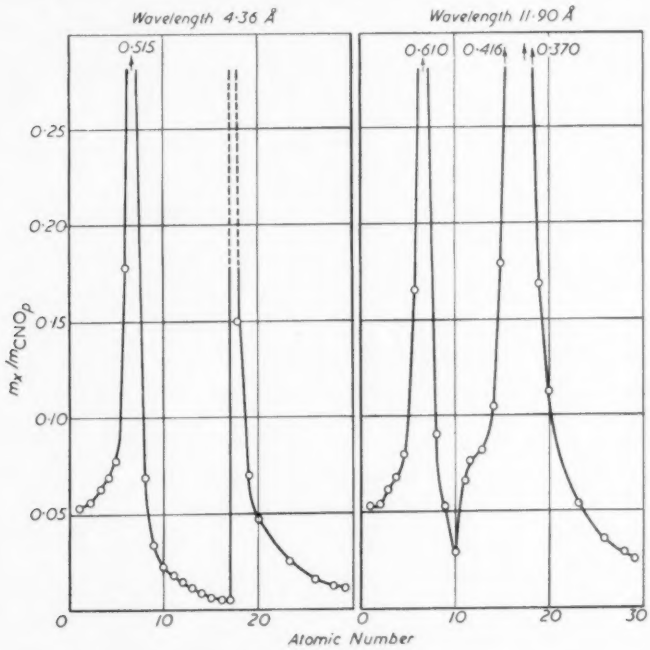


Figure 5. The quotient $\frac{m_x}{m_{CNOp}}$ as a function of the atomic number calculated for wavelengths 4·36 and 11·90 Å

From *Figure 5* it can be seen that there are two extreme ranges. One is situated at the C,N,O region. The other extreme range lies at the atomic number 18 and broadens at longer wavelengths. The increase of breadth is most prominent on the side towards lower atomic numbers. This depends on the influence of the K-absorption edge. In the parts of the curves where the derivative is positive, the systematic error is negative and *vice versa*. This fact indicates that the systematic errors can compensate each other, at least to some extent. It can also be seen from the figure that the systematic errors decrease if long wavelength x-rays are used. Calculations show that a continuous x-ray spectrum generated at 3,000 volts DC and filtered through 9 μ

aluminium is suitable for most biological material. For such x-rays about three quarters of the total intensity lies in the wavelength region 8 to 12 Å.

Taking into consideration both the intensity distribution in the continuous filtered x-ray spectrum and the different values of the allowed percentages *i.e.* m_x/m_{CNO_p}, the final concentrations for $\epsilon = 5$ per cent for the elements of biological interest can be calculated. The values are graphically given in per cent of m_{CNO_p} in *Figure 6*. For details concerning the calculations the reader is referred to the paper quoted above.

From *Figure 6* it is obvious that relatively large variations in the composition of C,N and O do not affect the result to any considerable extent (a small error in the mass determination). Of the other elements found in biological tissues only hydrogen might give a systematic error ϵ exceeding 5 per cent. The sum of the systematic errors introduced by the other elements can be shown not to exceed 5 per cent, but a correction factor for the protein hydrogen must be introduced. The error from the hydrogen in other compounds in the dried specimen is almost compensated by the errors from the trace elements. The reference system must consist mainly of carbon, nitrogen, and oxygen and is made of nitrated cellulose of known chemical composition and known mass per unit area.

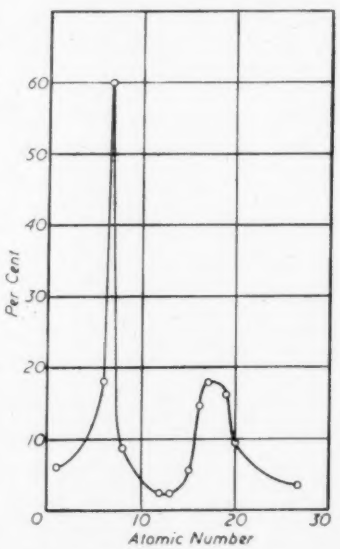


Figure 6. Maximal allowed percentage of m_x in relation to m_{CNO_p} that gives an error of 5 per cent in the mass determination, for 3,000 v X-rays filtered through 9 μ aluminium

Taking all corrections into consideration we get the following expression for the calculation of the mass of a cell structure

$$\frac{m_\alpha}{K_{Hp}} = \frac{K_{Href}}{K_{Hp}} \cdot m_{ref} \cdot \frac{(\mu/\rho)_{CNOref}}{(\mu/\rho)_{CNOp}} \qquad \ldots\ldots(12)$$

The symbols have the following meaning :

$\frac{m_\alpha}{K_{Hp}}$ = the weight of the biological structure under investigation

m_{ref} = the mass of the reference system having the same absorption as the sample

K_{Hp} = the correction factor for hydrogen in protein and has the value 0·925

K_{Href} = the correction factor for the hydrogen in the reference system and can be calculated when the chemical composition of the substance in the reference system is known

$(\mu/\rho)_{CNOp}$ = the mass absorption coefficient for C,N and O in protein as defined above

$(\mu/\rho)_{CNOref}$ = the mass absorption coefficient for C,N and O in the reference system.

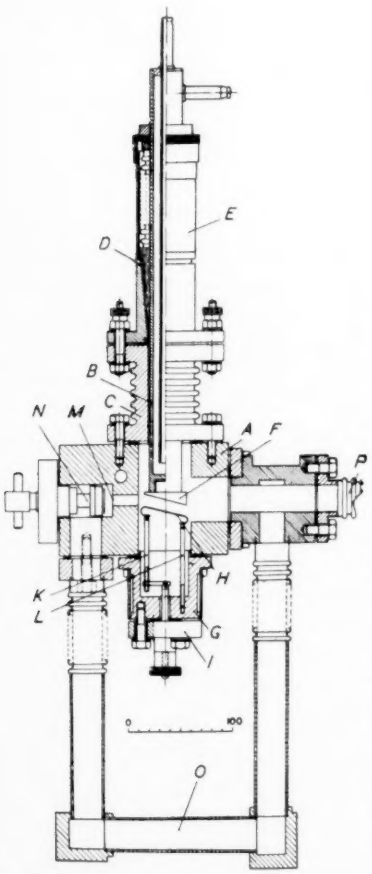


Figure 7a. Diagram of x-ray tube for production of very soft x-rays, used for determination of the mass of cell structures

This brief and necessarily incomplete theoretical discussion shows that it is possible to determine the mass of small biological objects by absorption measurements with long wavelength x-rays. Naturally the method is not applicable to tissues where high concentrations of other elements than carbon, nitrogen and oxygen occur *e.g.* bone tissue and chitin-like substances. For many biological problems it is enough to know the relative distribution of mass. The relative mass is obtained directly by comparison between the absorption in the cell structure and the absorption in the reference system. The mass of the cell structures thus is obtained in units of the reference system, and if the same reference system is used in the experiments the different values are directly comparable.

Apparatus— For the production of long wavelength x-rays a special x-ray tube must be used. The specimens must be exposed in high vacuum in order to minimize absorption by air. As extremely fine grained film emulsions are used the output of soft x-rays must be very high. A drawing of a suitable tube[34] is shown in *Figure* 7a.

The body of the tube *A* is made of forged brass. The anode *B* is isolated by a porcelain cylinder *C* from *A*. By the cone *D* and the bellows *E* the anode with the target *F* can be adjusted in high vacuum. The cathode holder *G* with the insulated part *I* supports the leads *K* and *L* to the filament *H*. A tungsten filament about 1 mm in diameter is used. The x-rays leave the inner part of the tube through the aperture *M* which is covered with a foil of aluminium 9 μ thick. The foil serves as a filter and protects the photographic film from light. The sample is mounted in the holder *N* which is removable. *O* is a vacuum connection between the central part of the tube and the sample space. The whole tube is evacuated through *P*. The anode and the body of the tube are water cooled. The joints are made vacuum tight by rubber gaskets. A photograph of the experimental arrangement is

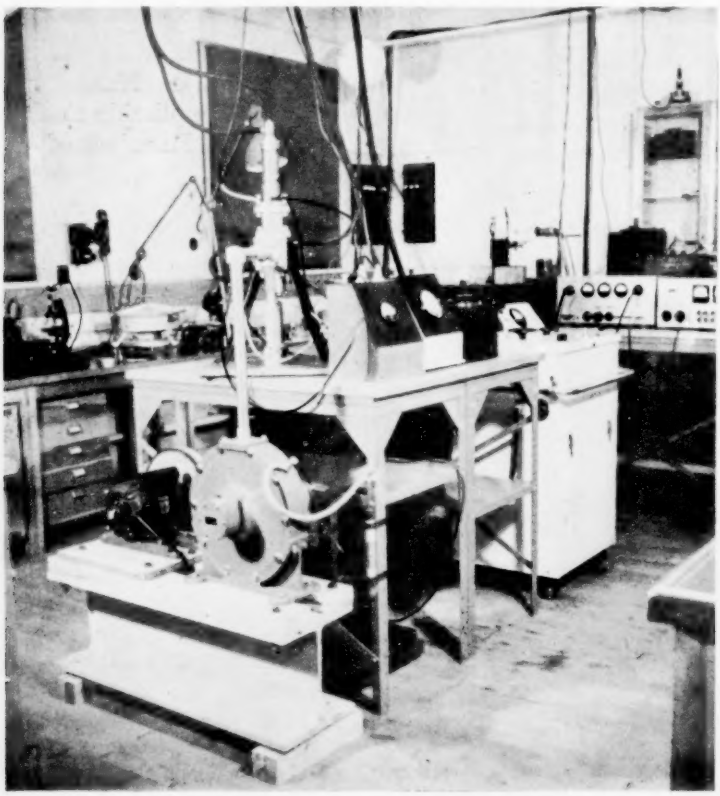


Figure 7b. *Photograph of the equipment used for the determination of the mass of cell structures: the x-ray tube stands on the table, with the molecular pump and forepump in front and the high voltage equipment, electronic voltage stabilizor and auxiliary instruments to the right*

shown in *Figure* 7b. The tube is evacuated with a molecular pump, backed by an ordinary fore-pump. The voltage, emission and vacuum are electronically controlled and regulated.

Technique— A careful mounting of the specimen, which has to be investigated in high vacuum, is important. Over the slit in the preparation holder A (*cf Figure 4*) a thin collodion foil is laid. The foil generally has a thickness of about 0·3 μ and serves as a supporting membrane for the biological preparation. On one half of the slit B is placed a smear or a section of the tissue under examination. On the other half of the membrane the reference system is laid, in the form of a step wedge. The x-ray absorption of the reference system should be of the same order as that of the sample. As the reference system should be built up mainly of carbon, nitrogen and oxygen thin collodion foils are used. The preparation is laid against the emulsion of a Lippman film and very good contact secured in order to minimize geometrical unsharpness. The preparation is then exposed with soft x-rays the voltage and filtration of which are chosen in accordance with the calculations given above. Slit illumination must, of course, be checked for evenness before the exposure is made. After development and fixation the microradiogram is washed in clean water and dried in a dust free atmosphere.

The microradiogram of the sample and wedge is then enlarged up to 500 times by photomicrography. The granularity of the Lippman film does not disturb photometric measurements at these magnifications. The different parts of the microradiogram and the wedge are photographed separately on a plate with constant exposure time, illumination and focusing (*Figure 8*). The part of the wedge shown in the figures is laid in the form of a cross in order to get the image of two foils in a small area. By this arrangement it is not necessary to know the shape of the density curve either for the Lippman film or for the photographic plate when performing the subsequent photometry. The density in the image of different cell structures is then compared with the density in the image of the wedge and the relative mass of the sample is obtained in units of the wedge. For absolute determination of the mass the calculation is performed in

Figure 8. Photomicrograph of a microradiogram showing a part of a muscle fibre and the reference system

In some photomicrographs of the microradiograms (*Figures 8, 10, 11, 14, 15, 16*), the white areas are absorbing x-rays. The microradiograms were registered with 3,000 v x-rays filtered through 9μ aluminium, except *Figure 5* which was registered with 5,000 volts filtered through 9μ aluminium.

accordance with formula 12. The resolving power for a specimen with good contrast is as small as 1 μ with this technique.

Applications— In this section a few applications of the technique for mass determination will be briefly described. For details the reader is referred to the original papers.

In *Figure 9* is shown a photomicrograph of a microradiogram of single nerve fibres. Single nerve fibres were obtained by microdissection of the ischiadic nerve from frog and the fibres dried at low temperature. During dissection before drying checks were made that the fibres were excitable and the salts in the fluids used for the dissection were washed away. From the picture it is obvious that the fibres have a tube-like structure, where the outer layer of the tube has a greater opacity to x-rays than has the inner portion. This outer layer seems to correspond histologically to the myelin sheath. The nodes of Ranvier can be seen and in the neighbourhood of each node the outer layer, presumably the myelin sheath, is thicker.

Figure 9. Photomicrograph of a microradiogram of single nerve fibres, obtained by microdissection of the ischiadic nerve from frog. In the picture the black areas are absorbing x-rays. The tubular structure is clear and inside the tubes there are inhomogeneities due to the shrinkage of the axoplasm during drying. In the neighbourhood of the nodes of Ranvier, the outer, stronger absorbing layer becomes thicker. Magnification c 200 × 200

Quantitative measurements on single nerve fibres indicate that the outer layer of the tube has a mass of about 0·3 to 0·4 $\times 10^{-12}$ gm μ^3, and the inner portion, presumably the axis cylinder, has five to eight times smaller mass. The nerve fibre is thus a rigid tube with most of its mass contributed by the myelin sheath (*see Figure 10*). If the specific density of the

Figure 10. Frozen section through a bundle of nerve fibres from the ischiadic nerve of frog. The cross section is rather thick but in some places the high absorbing power of the myelin sheath and the almost nonabsorbing axis cylinder can be seen. Magnification c 100 × 100

axis cylinder is assumed to be 1 in the living fibre the measurements obtained indicate that the water content of the axoplasm is 90 to 95 per cent.

Fibres, the mass of which had been determined, were extracted with lipid solvents. After extraction a repeat determination of mass was performed. The lipid content thus was obtained by the difference between these two measurements. After extraction the myelin sheath was not homogeneous as before. The middle third of the sheath had lost the most material. In this middle portion about half of the mass was extracted by lipid solvents. No appreciable loss of mass was found in the axis cylinder.

Living single nerve fibres were electrically stimulated. Changes in the mass of the axis cylinder close to the nodes were observed but this problem demands detailed study. A detailed report on the examination of nerve fibres is found in the paper by ENGSTRÖM and LÜTHY[36].

In striated muscle fibres the cross striations come out clearly. *Figure 11* shows a microradiograph of a section through a muscle from a *Chironomus* larva. Also it can be seen that the mass is not equally distributed within the bands. The bands with a high opacity to x-rays seem to correspond to the anisotropic segments. When the mass in the different bands was determined it was found that the ratio between the mass of the strong and weak absorbing segments is about 1·8 : 1. Consulting *Figure 6* it is obvious that this represents a true difference in mass and is not an effect of elements with higher atomic number *e.g.* potassium. A fuller account will be published later.

Figure 11. Section through a muscle fibre from Chironomus. *The cross bands can be clearly seen. The contrast in the upper end of the fibre is good but decreases towards the other end, since the thickness of the section is not the same over the whole fibre. In such a small area as required for measurement of the difference in mass between two neighbouring bands the thickness can be considered to be the same. Magnification c* 300 × 300

The distribution of mass has been studied in animal and plant cells. In most animal cells the nucleus seems to have a lower mass per unit volume than the cytoplasm or the nucleolus (*Figure 12*). In certain plant cells *e.g.* growing cells in the onion root tip, the nucleus has a greater density than the cytoplasm (*Figure 13*). The cytoplasm of nerve cells from the spinal cord has about double the mass of the nucleus per

unit volume. The nucleolus has at least the same mass per unit volume as the cytoplasm[35].

Figure 12. Photomicrograph of a microradiogram of a nerve cell in a microtrome section from spinal cord, using 3,000 v radiation filtered through 9μ aluminium. The white areas are absorbing x-rays. The strong x-ray absorption of the nucleolus, indicating a high density of material, is prominent. The nucleus has less x-ray absorption than the cytoplasm. Magnification c 300 × 300

Figure 13. Longitudinal section through an onion root tip. The microradiogram registered as in Figure 2. The nuclei give strong contrast in comparison with the almost empty cytoplasm. Magnification 60 × 60

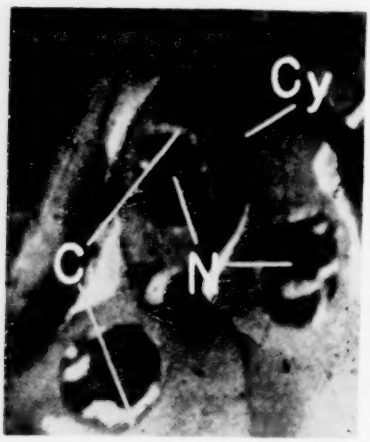


Figure 14. Microtome section from the salivary gland of Chironomus. *Note the low mass per unit area of the nucleus* N *in comparison with that of the cytoplasm* Cy. *Within the nucleus the strong x-ray absorbing chromosomes* C *can be seen. Magnification c 100 × 100*

Figure 14 illustrates the difference in mass between nucleus and cytoplasm in cells from the salivary gland of *Chironomus*. The high density of the chromosomes C in the nucleus N is striking. In the salivary gland chromosomes there are great differences in mass between the different bands as illustrated in *Figure 15*.

Finally in *Figure 16* is shown a photomicrograph of a section of skin. The amount of substance was determined in different layers and details may be obtained in the paper quoted above[34, 35].

Conclusions— By the measurement of a broad band of x-rays, properly selected, it is possible to determine the mass of cellular structures as small as 1 μ in diameter. The rela-

Figure 15. Giant chromosome from Chironomus *obtained by microdissection. The picture is somewhat unsharp due to the thickness of the chromosome. This not too good picture, however, shows the great difference in mass between the different groups of bands. Magnification c* 500×500

tive mass is obtained directly by comparison with a reference system and the absolute mass can be computed. For most biological problems, however, it is only necessary to know the relative distribution of mass. By determining the mass of a structure and then extracting the specimen or treating it with enzymes a repeat determination of mass gives information about groups of substances specifically removed.

The introduction of elements with high atomic numbers which react stoichiometrically with certain substances in the sample makes it possible to estimate the amount of these substances by the increase in x-ray absorption.

Figure 16. A microtome section (5μ) through skin epithelium. Stratum corneum has a strong x-ray absorption, thus a high mass. The underlying epithelial cells have a smaller mass per unit area. In this layer the single cells with their nuclei can be seen at higher magnification. The collagenous fibres are clearly seen in the subcutaneus tissue. Magnification c 50×50

QUANTITATIVE DETERMINATION OF ELEMENTS IN BIOLOGICAL MATERIAL BY X-RAY ABSORPTION SPECTROGRAPHY

In the x-ray absorption spectrum of the elements absorption discontinuities, sometimes called absorption edges or absorption jumps, appear at wavelengths characteristic for each element. The positions of these edges in the x-ray absorption spectrum are an atomic property of the element, thus they are not affected by the chemical state. Such a series of discontinuities is shown schematically in *Figure 17*. The shape of an x-ray absorption curve for an element is plotted for a wide wavelength range. The designation of the absorption edges appears in the figure. The elements with high atomic numbers have their absorption edges at very short wavelengths and the position in the spectrum of the absorption edge is shifted towards longer wavelengths with decreasing atomic number.

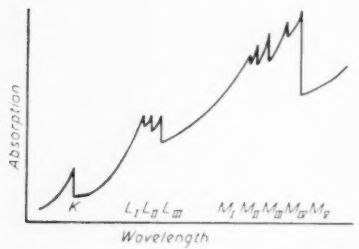


Figure 17. Schematic representation of the x-ray absorption spectrum for an element

Table II. Critical Absorption Wavelengths[2, 6] *for some Elements of Biological Interest*

Element	*K-absorption edge* Å	*Element*	*K-absorption edge* Å	*Element*	L_{III}*-absorption edge* Å
6 C	43·5	16 S	5·01	20 Ca	35·6
7 N	31·8	17 Cl	4·38	29 Cu	13·2
8 O	23·5	19 K	3·43	47 Ag	3·69
9 F	18·0	20 Ca	3·06	53 I	2·71
12 Mg	9·50	26 Fe	1·74	79 Au	1·04
15 P	5·77	29 Cu	1·38	80 Hg	1·01

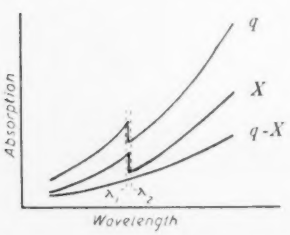


Figure 18. Absorption of a compound sample q divided into two fractions one comprising the element sought X and the second the other elements (q−X)

Table II gives the wavelength values for some elements of biological interest. It is clear that the *K*-absorption edges for the elements of major biological interest, C,N,O,Na,P,S,Cl, K,Ca and Fe lie within a wavelength region which is favourable for the measurement of absorptions in a biological specimen.

In a sample containing several elements the estimation of a particular element is made by measuring the absorption of x-rays in the vicinity of an absorption edge

of that element[37-40]. Let the mass of the whole sample be q gm/sq cm (*Figure 18*). The absorption for the sample can be broken up into two components, one comprising the element being determined x and one comprising all the other elements. The absorption of x-rays of the wavelengths λ_1 and λ_2 per sq cm of sample is, according to equation 2

$$i_1 = I_1 \exp\left\{ -\frac{\mu_1}{\rho} X + \frac{\mu_1'}{\rho'}(q - X) \right\}$$
$$i_2 = I_2 \exp\left\{ -\frac{\mu_2}{\rho} X + \frac{\mu_2'}{\rho'}(q - X) \right\} \quad \ldots\ldots(13)$$

The symbols have the following meaning :

λ_1 = wavelength on short wave side of absorption edge
λ_2 = wavelength on long wave side of absorption edge
X = mass of sought-for element in gm/sq cm
q = mass of whole sample in gm/sq cm
μ_1/ρ = mass absorption coefficient for X in the wavelength λ_1
μ_2/ρ = mass absorption coefficient for X in the wavelength λ_2
μ'_1/ρ' = hypothetical mass absorption coefficient for $(q - X)$ in wavelength λ_1
μ_2'/ρ' = hypothetical mass absorption coefficient for $(q - X)$ in wavelength λ_2
I_1 = intensity of the incident radiation in wavelength λ_1
I_2 = intensity of the incident radiation in wavelength λ_2
i_1 = intensity of transmitted radiation in wavelength λ_1
i_2 = intensity of transmitted radiation in wavelength λ_2

If $(q - X)$ is eliminated the following expression is obtained

$$X = \frac{\ln\left(\frac{i_2}{I_2}\right) \cdot \frac{\mu_1'}{\mu_2'} - \ln\left(\frac{i_1}{I_1}\right)}{\frac{\mu_1}{\rho} - \frac{\mu_2}{\rho} \cdot \frac{\mu_1'}{\mu_2'}} \quad \ldots\ldots(14)$$

As has been said before the scattering of x-rays can be ignored in comparison with the photoelectric absorption in the wavelength region referred to here. From equations 13 and 14 it can be seen, however, that even if the scattering were appreciable it would not affect the analytical result. The scattering effect for the sample is the same both for λ_1 and λ_2 as those two wavelengths lie close to each other.

The quotients i_1/I_1 and i_2/I_2 in equation 14 can be determined experimentally and the values of the mass absorption coefficients can be taken from text books in x-ray physics or standard tables. In order to calculate X it is necessary to determine the quotient between the two absorption coefficients for the foreign substances in the sample *i.e.* the absorption coefficients for $(q - X)$. If the wavelengths λ_1 and λ_2 both

lie quite close to the absorption edge, the quotient μ_1'/μ_2' can be replaced by $(\lambda_1/\lambda_2)^p$ which can be calculated from equations 2 and 6. If the area analysed is S sq cm the amount of the element being determined X is

$$X = \frac{\ln\left(\frac{i_2}{I_2}\right) \cdot \left(\frac{\lambda_1}{\lambda_2}\right)^p - \ln\left(\frac{i_1}{I_1}\right)}{\frac{\mu_1}{\rho} - \frac{\mu_2}{\rho} \cdot \left(\frac{\lambda_1}{\lambda_2}\right)^p} \cdot S \qquad \ldots.(15)$$

The value of the exponent p is discussed in connection with equation 7.

When the wavelengths λ_1 and λ_2 are selected quite close to the edge and the intensity of the incident radiation is assumed to be the same in the two wavelengths ($I_1 = I_2$), the following simplified equation is obtained for the calculation of X

$$i_1 = i_2 \cdot \exp\left\{-X\left(\frac{\mu_1}{\rho} - \frac{\mu_2}{\rho}\right)\right\} \qquad \ldots.(16)$$

The limit of sensitivity for the method can be calculated from equation 16 when the accuracy of the absorption determination is known. When λ_1 and λ_2 are not situated close to the absorption edge the exponent p must be determined with greater accuracy than when the two wavelengths lie close to the edge. In such cases it is preferable to determine p experimentally on substances of known composition.

The ideal method, however, is to register a continuous absorption spectrum in the neighbourhood of an absorption edge and calculate the amount sought for according to equation 16. In this instance the height of the absorption jump is obtained directly without introducing any correction for the selection of wavelengths. This technique is possible only when a continuous x-ray spectrum is available and it is difficult to combine with microradiography as a very great number of microradiograms must be taken to approach a continuous spectrum.

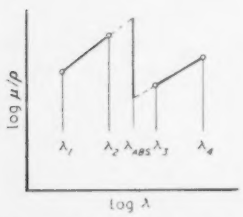


Figure 19. Mass absorption coefficient as a function of the wavelength in logarithmic scale

The height of the absorption jump can be extrapolated in another way. If we take the logarithm of both sides of equation 6, ignoring the scattering we obtain

$$\log \frac{\mu}{\rho} = n \cdot \log \lambda \qquad \ldots.(17)$$

If a curve is drawn with $\log (\mu/\rho)$ as a function of $\log \lambda$ we obtain two straight lines as shown in *Figure 19*. The inclination on the two sides of the absorption jump is slightly different depending on the value of n.

In the sample the absorption, i/I, for different wavelengths $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ . . . is determined experimentally. If equations 1 and 17 are combined it can be shown that the logarithm of ln (I/i) is a straight line function of log λ. The height of the absorption jump thus can be extrapolated in a double logarithmic system where ln (I/i) is plotted against λ (for details see ENGSTRÖM[38]). Knowing the height of the absorption jump the amount of the element sought for can be calculated from equation 16. This time consuming procedure for determination of an element can be used in most cases even when there are large quantities of foreign elements $(q - X)$ present in the sample.

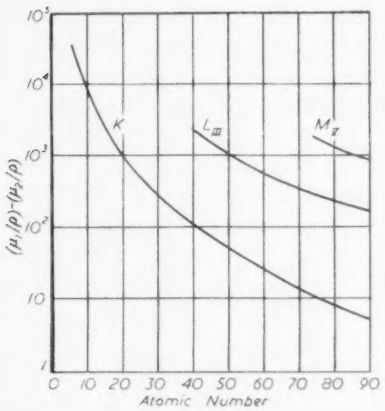


Figure 20. Difference of the mass absorption coefficients on both sides of the absorption edge, $\left(\frac{\mu_1}{\rho_1} - \frac{\mu_2}{\rho_1}\right)$, as a function of the atomic number. The calculations refer to K, L_{III} and M_V edges

Sensitivity of the method— The next step in the calculations is to investigate the sensitivity of the x-ray absorption method and judge its applicability for solving problems in the chemistry of cells and tissues. From equation 16 it can be seen that small amounts of an element X can be determined when $(\mu_1/\rho) - (\mu_2/\rho)$ has a high value. The numerical value of this difference as already stated increases with decreasing atomic number. In *Figure* 20 is shown a graph where $(\mu_1/\rho) - (\mu_2/\rho)$ is plotted as a function of the atomic number. The increase in the value of $(\mu_1/\rho) - (\mu_2/\rho)$ depends on two factors. One is that the absorption edges for elements with low atomic numbers are in a region where the x-ray absorption is strong, thus giving a large value for the mass absorption coefficients. The other is that the relative size of the absorption jump increases with decreasing atomic number.

The smallest determinable amount of an element depends on the accuracy with which the quotients between the incident and transmitted radiations can be determined, and also on the accuracy demanded in the determination of X. If in equation 16 an intensity difference of 5 per cent between i_1 and i_2 is assumed to be the smallest measurable the amount of different elements corresponding to such a difference can be calculated. In *Figure 21* the calculation is presented. In this figure the calculations are performed only for the K-absorption edge, as this edge is mainly used for analysis of biological material.

If it is possible to determine experimentally the absorption of x-rays in the sample with great accuracy still smaller amounts can be estimated. The diagram in *Figure 21* can be used to indicate proper layer thicknesses for the determination of certain elements in a biological tissue. When photographic-photometric methods are used to measure the absorption of x-rays the minimal amount of the element being determined per unit area must be about 5 times larger than the figure obtained from the diagram if X is to be determined with an accuracy of 10 to 15 per cent. It is then assumed that the photographic-photometric procedure for measuring the quotient between the intensities of the incident and transmitted radiations can be performed with an accuracy of a few per cent.

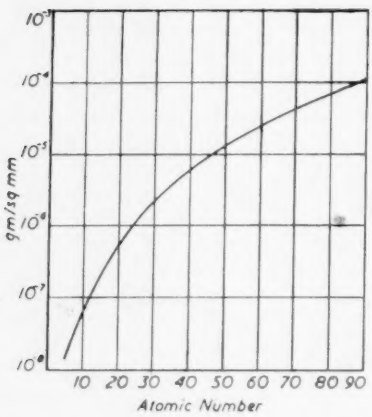


Figure 21. Smallest determinable amount of an element

Thus it is possible to determine quantitatively several elements in biological tissues of thicknesses corresponding to a single cell layer. This holds good for elements occurring in relatively large concentrations. For elements occurring in minute concentrations the layer thickness must be greater, and the absorption of x-rays must also be determined with high accuracy.

Technique and apparatus— The sample for analysis is prepared and mounted as for mass-determination. The preparation is laid on one half of the slit of the sample holder, the reference system on the other half (*Figure 4*). When electrical methods are used to determine the absorption of x-rays no reference system is needed. The reference system must be used when soft x-rays are employed to obtain a photographic image of the sample. The density response of the photographic emulsion when soft x-rays are used is not linear with the incident x-ray energy as is true for hard x-rays. Only for low densities is there a proportionality to the intensity of the incident x-rays[41].

With monochromatic x-rays microradiograms of the sample are registered. The wavelengths selected depend on the element being determined. It is preferable to use the strongest characteristic x-ray emission lines as they have a relatively high intensity. Usually the $K\alpha$ or $L\alpha$ radiation is used. Proper emission lines can be selected from text books in x-ray physics.

Line radiation is obtained by primary or secondary excitation. In

the first instance the material, the line radiation of which is desired, is soldered on to or rubbed into the anode of the x-ray tube. Thus the electrons impinge on the material. When the characteristic x-rays (line radiation) are secondarily excited primary x-rays strike the material which is to emit x-rays. The primary radiation must in this case have wavelengths shorter than the absorption edge for the particular group of lines desired. The yield of secondary x-rays rapidly decreases with decreasing atomic number of the secondary radiator, which is unfavourable for the investigation of biological material.

In case the lines are primarily excited the total voltage on the x-ray tube usually must not exceed the double value for the voltage corresponding to excitation of first order as these hard x-rays are reflected as higher orders at the same Bragg angle. The primary-excited x-rays leaving the tube consist of a continuous and a superimposed line spectrum. The particular x-ray spectral line is isolated by reflection in a crystal grating. The microradiogram of the sample and wedge is registered with such a spectrally pure radiation.

When the x-rays are secondarily excited it is most favourable to purify them by reflection in a crystal. However, it is also possible to filter such radiation properly. In a *K*-emission spectrum all lines except the one desired can be depressed by proper filtering. The primary x-rays also must be excited in such a way that it is possible to filter off the primary rays scattered by the secondary radiator[42, 43].

For the isolation of the desired x-ray emission lines an ordinary x-ray vacuum spectrograph *e.g.* one of Siegbahn's design, can be used. As the films used for taking microradiograms of the sample are very insensitive, because of the very small grain size, the exposure times are very long. This disadvantage can be eliminated partially by using high luminosity spectrographs with curved crystal gratings. The difficulties encountered in obtaining monochromatic x-rays are much increased in the extremely soft x-ray region where the absorption edges of carbon, nitrogen and oxygen are situated. A new spectrograph for obtaining monochromatic x-rays in this region, utilizing a curved crystal grating, is now being constructed by the author.

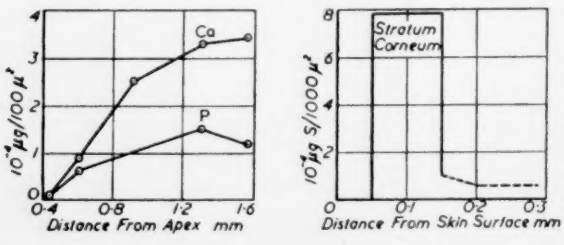
Applications— In order to check the method of chemical analysis by x-ray absorption spectroscopy determinations of elements in samples of known composition were performed. In microcuvettes of a few cu mm capacity the amounts of zinc and iron were determined. *Table III* gives the results from one experiment, where the amount of zinc was determined in solutions to which varying amounts of other elements had been added in order to find out the disturbing effect of the foreign elements. The cuvettes were made of aluminium foils of varying thickness. Zinc has its *K*-absorption edge at 1·28 Å. The $L\alpha$

Table III. Determination of Zinc in Microcuvettes

Composition of the solution mg/cu mm	*Zn taken* μg/sq mm *cuvette* 0·56(4)*mm*	$\lambda_1 = 1{\cdot}27$ Å		$\lambda_2 = 1{\cdot}31$ Å		*Zn found* μg/sq *mm*	*Deviation*	
		DI_1	Di_1	DI_2	Di_2		μg	*per cent*
0·04728 $ZnSO_4.7H_2O$	6·1	0·065	0·025	0·058	0·024	5·8	−0·3	−4·9
0·3377 $ZnSO_4.7H_2O$	43·3	0·168	0·033	0·218	0·106	39·0	−4·3	−9·9
0·4728 $ZnSO_4.7H_2O$	60·6	0·194	0·022	0·223	0·100	58·1	−2·5	−4·1
0·2360 $Zn(NO_3)_2.6H_2O$	29·2	0·234	0·068	0·233	0·129	28·0	−1·2	−4·5
0·2364 $ZnSO_4.7H_2O$ 0·0303 $CuSO_4.5H_2O$ 0·0220 $FeSO_4.7H_2O$	30·3	0·196	0·030	0·215	0·066	32·0	+1·7	+5·6
0·07880 $ZnSO_4.7H_2O$ 0·0202 $CuSO_4.5H_2O$ 0·0146 $FeSO_4.7H_2O$	10·1	0·079	0·031	0·131	0·063	10·5	+0·4	+4·0
0·1573 $Zn(NO_3)_2.6H_2O$ 0·0321 $BaBr_2.2H_2O$	19·5	0·133	0·035	0·152	0·060	19·4	−0·1	−0·5

lines from gold and platinum with the wavelengths 1·27 and 1·31 Å respectively were used for the determination of the height of the absorption jump. The measurement of the absorption in the small cuvettes was made by repeated photometric measurements in the photographic x-ray image of the cuvette. For the wavelengths used there is a direct proportionality between the density of the film and the intensity of the incident x-ray beam up to a density value of about 0·25. In the table, D indicates the density of the film, I, i, 1 and 2 have the same meaning as in formula 15. It can be seen that the analytical error is about 10 per cent.

Some histochemical applications are shown in *Figure 22*. In growing cells from guinea-pig teeth the amounts of calcium and phosphorus were determined. Calcium has its K-absorption edge at 3·06 Å and phosphorus has the edge at 5·77 Å. For the determination of calcium the $K\alpha$ lines from Sc (3·03 Å) and Ca (3·35 Å) were used and for the determination of phosphorus the P $K\alpha$ and S $K\alpha$ lines (6·14 and 5·36 Å respectively). From the figure it can be seen that the content of



a b

Figure 22. Determination of a *calcium and phosphorus in dentine cells, and* b *of sulphur in different layers in the skin* [38, 43]

these two elements increases rapidly with the distance from the apex of the tooth. The analytical findings can be correlated with the growth and differentiation of the dentine cells.

In *Figure 16* a microradiograph of a section of skin is shown. This microradiogram is registered with polychromatic x-rays generated at 3,000 volts and filtered through 9 μ aluminium. If the microradiograms are taken with the $K\alpha$ lines from sulphur and chlorine (5·36 and 4·72 Å respectively) which lie on either side of the K-absorption edge for sulphur (5·01 Å) the amount of the latter element can be determined. In *Figure 22*b such a determination has been performed. It can be seen that the content of sulphur is highest in the stratum corneum of the epithelium and decreases towards the inner layers. These analytical results, also, can be correlated with the growth and differentiation of the cells in the skin epithelium. As the measurements of the x-ray absorption for the two types of tissue shown above are correct within a few per cent, the values in *Figures 22*a and b have an error of about 15 per cent.

Conclusions— The discussion presented above shows that several elements in a biological tissue can be determined within a volume corresponding to that of a single mammalian cell. Knowing the average approximate chemical composition of a tissue, the curves and calculations given in this section make it possible to estimate the possibility of analysis and the proper thickness of the specimen to be analysed in each case.

X-RAY EMISSION ANALYSIS OF BIOLOGICAL MATERIAL

When performing an x-ray emission analysis of a substance the characteristic x-ray emission lines of the sample are primarily or secondarily excited (*see* p 188). The elements of which the radiator is composed can be determined by estimating the wavelengths of the emitted radiation. Intensity measurements in the line spectrum make it possible to determine the amounts of the elements. As biological material to a very large extent consists of organic substances which would decompose and evaporate under electron bombardment in an x-ray tube the method of secondary excitation is the most suitable. Also, from the point of view that the area analysed must be carefully localized the method of secondary excitation has advantages. As x-ray emission methods up to date have found little application to the analysis of biological material this section is very condensed and only the most promising methods are mentioned.

In general, the elements to be determined in the biological material have low atomic numbers. This is very unfavourable as the yield of secondary x-rays rapidly decreases with atomic number. A rough estimation of the intensity conditions indicates that the emission

spectroscopy of biological material can be used only under certain special conditions. The combination of a very small intensity of the secondary x-rays from a biological sample with the necessity to be able to localize the area analysed indicates that the best experimental arrangement would be a curved crystal in 'image position' in the spectrograph.

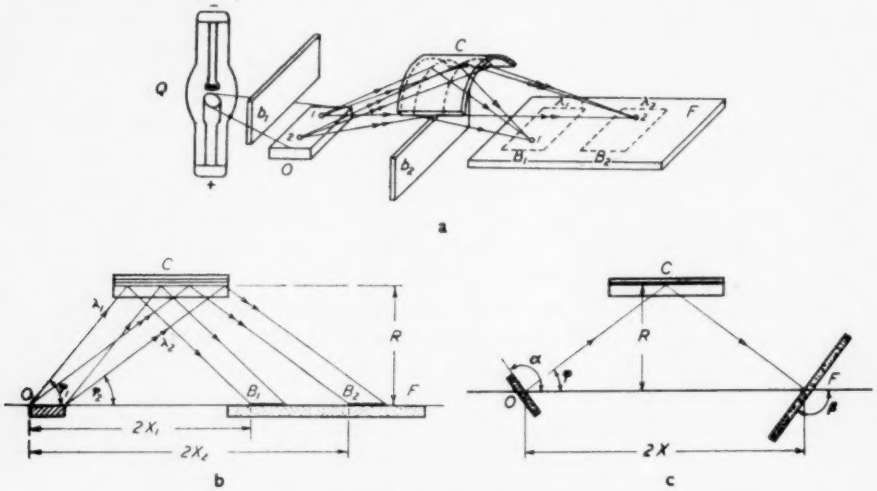


Figure 23. X-ray image spectrograph[44]. a *General arrangement,* b *Position of the image in relation to the sample,* c *Arrangement for highest resolving power*

*Figure 23*a illustrates the principle of the x-ray image spectrograph, developed by VON HAMOS[44-46]. In the x-ray tube *Q* primary x-rays of proper wavelengths are generated. These x-rays induce characteristic x-rays in the sample *O*. The secondary characteristic x-rays are reflected according to the Bragg law in the cylindrical crystal *C* and produce an image on the photographic film *F*. The crystal has the radius *R*. The characteristic radiation of each element, λ_1 and λ_2, will form images B_1 and B_2 at certain places on the film, as illustrated in the figure. Thus the position of the image on the film gives information about the nature of the element and the intensity of the radiation (density of the image), has a relationship to the amount of that element. *Figure 23*b shows a longitudinal section of the image spectrograph. The distance $2X$ between corresponding points in the sample and in the image is given by

$$2X = 2R \cot \varphi \qquad \ldots.(18)$$

and, introducing the Bragg formula

$$n\lambda = 2d \sin \varphi \qquad \ldots.(19)$$

we get

$$2X = 2R\sqrt{\left(\frac{2d}{n\lambda}\right)^2 - 1} \qquad \ldots\ldots(20)$$

In the formula d is the grating constant of the crystal and n the order of reflection. The geometrical properties of the image spectrograph have been discussed by VON HAMOS[44-46] in detail. By changing the angles between the sample plane and the plane of the film relative to the longitudinal axis for the crystal curvature, it is possible to obtain an enlarged image of the sample. The image, however, is distorted. The highest resolving power of the system is obtained when the plane of the sample and the film are arranged in accordance with *Figure 23c*. Unit magnification is obtained and the angles that the film and sample make with the longitudinal axis are

$$\alpha = \varphi + 90^\circ \qquad \beta = -\varphi - 90^\circ$$

The quality of the image is good and if the grain of the film is not too coarse, microscopic examination of the film is possible. The best image quality is obtained when the object is situated near the longitudinal axis of curvature of the crystal and the resolving power of the crystal mirror is not limited by geometric aberration as much as by chromatic aberration due to the nature of x-rays. The order of magnitude of the chromatic aberration is 1/1,000 of the crystal radius[43].

HOPPE and TRURNIT[47] published a paper on the theory of spherical crystal lenses for monochromatic x-rays. From their calculations it can be seen that a symmetrical rotational surface, coated with a crystalline material, theoretically might give a maximal resolving power of about 1 μ.

When a thick object such as a metal block is irradiated with primary x-rays and the intensity of the secondary x-rays is measured the concentration of a particular element can be calculated. As the thickness of the object is decreased, the total intensity of secondary radiation remains constant until the absorption of the secondary radiation generated at the maximal depth in the sample is no longer complete (see *Figure 24*). For very thin samples where the absorption of the primary and secondary x-rays in the sample itself is negligible, the intensity of the characteristic secondary radiation is proportional to the mass per unit area of the element emitting the radiation. Taking these calculations into consideration the method of emission analysis would be suitable for determination of elements of

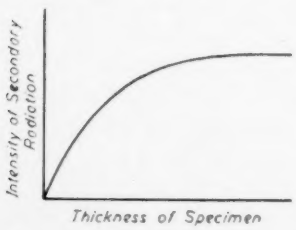


Figure 24. Intensity of secondary line radiation as a function of sample thickness[16]

atomic numbers higher than those usually found in biological tissues. The exposure times, however, are very long as the concentrations of these high atomic elements are small and thin sections have to be used.

The method of image spectroscopy has proved to be useful in metallurgy and VON HAMOS and ENGSTRÖM[48] tried to use it for the determination of zinc and iron in sections of biological tissues. The exposure times were very long and the images obtained did not allow a high magnification. It might be expected, however, that the method will find applications in the analysis of biological material with relatively high concentration of elements with atomic number above 15 *e.g.* bone tissue.

The methods described may however be useful to make an image of a sample with monochromatic x-rays. Instead of exciting the sample to send out its characteristic x-rays, monochromatic x-rays are allowed to pass through the sample and the absorption image is recorded. A great difficulty is the necessity of working with extremely soft x-rays when analysing biological material. In the extremely soft x-ray region as said before, the generation, monochromatization and measurement of radiation is very complicated.

CONCLUSIONS

The use of soft x-rays permits the quantitative analysis of small volumes of biological material. The utilization of x-ray absorption spectroscopy makes it possible to determine the total mass as well as the amounts of certain elements in a cell structure.

When determining the mass polychromatic x-rays are used. The mass determination is based on the assumption that the main constituents of the biological material are carbon, nitrogen and oxygen. An element with a high atomic number would thus introduce an error because its real contribution to total mass would be lower than that calculated from its x-ray absorption. Such errors, however, can be minimized by proper selection of the wavelengths in the polychromatic x-ray spectrum. Therefore it is necessary to calculate the proper distribution of wavelengths in the x-ray spectrum used for each type of material investigated.

For the determination of elements monochromatic x-rays are used. By the ordinary photographic-photometric technique certain elements can be determined in a volume corresponding to that of a single cell. If the accuracy of the technique for measuring the absorption of x-rays can be increased it seems possible that also other elements occurring in smaller concentrations can be determined. When very small amounts are determined, however, the error in the determination of

the particular element becomes greater, a fact which must be kept in mind when judging the results.

The use of x-ray emission spectroscopy seems at present not to be generally applicable for the analysis of biological material.

The preparation of the samples for analysis is complicated for the specimens must be investigated in vacuum. As it is impossible to investigate living cells and tissues great attention must be paid to the artifacts that may be introduced by fixing and drying procedures.

Once the technical difficulties are satisfactorily overcome and appropriate care is taken in the interpretation of data, the methods presented in this paper can be and have been profitably applied to many hitherto inaccessible problems in cell biology. Total mass can now be used as a basis upon which to express elementary analyses of cellular structures. The mass determination method can be extended also to include certain biochemical analyses of more general nature for example the determination of percentage of matter extractable or digestible by various biochemical agencies.

One may also determine the ratio between total amounts of two elements for the purpose of determining the physiological state of a cell or a tissue or eventually to assist in the identification of intracellular substances (*e.g.* nitrogen/phosphorus ratio). Finally there is now the possibility of investigating changes in the total mass of cell organelles during various natural processes such as growth and differentiation.

REFERENCES

[1] KLEIN, O. and NISHINA, Y. *Ztschr. f. Physik.* 52 (1928) 853
[2] COMPTON, A. H. and ALLISON, S. K. *X-rays in Theory and Experiment* New York, 1943
[3] VICTOREEN, J. A. *J. Applied Phys.* 14 (1943) 95
[4] — *ibid* 19 (1948) 855
[5] JÖNSSON, E. *Absorptionsmessungen im Langwelligen Röntgengebiet Thesis* Uppsala, 1928
[6] SIEGBAHN, M. *Spektroskopie der Röntgenstrahlen* Berlin, 1931
[7] KIRKPATRICK, P. and BAEZ, A. V. *J. opt. Soc. Amer.* 38 (1948) 66
[8] — — *ibid* 39 (1949) 96
[9] HAYCOCK, C. T. and NEVILLE, F. H. *Trans. chem. Soc.* 73 (1898) 714
[10] GOBY, P. *C.R. Acad. Sci., Paris* 156 (1913) 686
[11] DAUVILLIER, A. *ibid* 190 (1930) 1287
[12] LAMARQUE, P. *ibid* 202 (1936) 684
[13] — *J. Radiol. et Electr.* 20 (1936) 325
[14] — *Pr. Méd.* 44 (1936) 478
[15] — *Bull. Hist. Appl.* 14 (1937) 5
[16] — *J. Belge Radiol.* 27 (1938) 109
[17] TURCHINI, J. *Bull. Hist. Appl.* 14 (1937) 17
[18] — *Schweiz. Ztschr. Path. Bakt.* 5 (1942) 10
[19] CLARK, G L. *Applied X-rays* New York, 1940
[20] — In *Colloid Chemistry* vol. 5 Ed. by A. E. ALEXANDER New York, 1944
[21] — *Radiology* 49 (1947) 483
[22] MADDIGAN, S. E. *Industr. Radiography* 4 (1946) nr 4
[23] SPROULL, W. T. *X-rays in Practice* New York, 1946
[24] SIEVERT, R. *Acta Radiol.* 17 (1936) 299
[25] BOHATYRTSCHUK, F. *ibid* 25 (1944) 351

[26] BARCLAY, A. E. and LEATHERDALE, D. *Brit. J. Radiol.* 21 (1948) 544
[27] — *ibid* 20 (1947) 394
[28] — *ibid* 22 (1949) 269
[29] ENGSTRÖM, A. *Acta Radiol.* 31 (1949) 504
[30] CROWTHER, J. A. *Handbook of Industrial Radiology, pp 151–182* London, 1946
[31] VON ARDENNE, M. *Elektronen-Übermikroskopie* Berlin, 1940
[32] TRILLAT, J. J. *J. Appl. Phys.* 19 (1948) 844
[33] SEEMAN, H. E. *ibid* 20 (1949) 231
[34] ENGSTRÖM, A. and LINDSTRÖM, B. *Biochim. et Biophys. Acta* 4 (1950) 351
[35] — — *Nature, Lond.* 163 (1949) 563
[36] — and LÜTHY, H. *Experientia* 5 (1949) 244
[37] — *Nature, Lond.* 158 (1946) 664
[38] — *Acta Radiol. Suppl.* 63 (1946)
[39] — *Biochim. et Biophys. Acta* 1 (1947) 428
[40] — Instruments and measurements conference Stockholm 1947 Transactions
[41] — and LINDSTRÖM, B. *Experientia* 3 (1947) nr 10
[42] — *Rev. Sci Instrum.* 18 (1947) 681
[43] — and LINDSTRÖM, B. *Experientia* 3 (1947) 191
[44] VON HAMOS, L. *J. Sci. Instrum.* 15 (1938) 87
[45] — *Z. Krist. A* 101 (1939) 17
[46] — *Ark. Mat. Astr. Fys.* 31 (1944/45) 1
[47] HOPPE, W. and TRURNIT, H. J. *Ztschr. Nat. forsch.* 2a (1947) 608
[48] VON HAMOS, L. and ENGSTRÖM, A. *Acta Radiol.* 25 (1944) 325

8

THE TOLERANCE OF MAN FOR RADIOACTIVE ISOTOPES

J. F. Loutit

THE development of the chain-reacting pile has led to the production of enormous quantities of radioactive fission products. Few, if any of these, have any practical immediate use, but they are necessary by-products of the process of fission of U^{235} or Pu^{239}. As the production of energy by nuclear fission is a process which has come to stay it has been necessary to consider the effect on the living things on this earth of the inevitable escape of a proportion of these radioactive isotopes into air and water.

As egoists our first consideration is for the future wellbeing of the human race, but because of our dependence for food and clothing on the plant and animal kingdoms these are not to be forgotten. Moreover man does not live on food alone. His most fundamental requirements are water and air. The purity of the atmosphere and water supplies is, therefore, the chief consideration. If these are uncontaminated, or only minimally contaminated, it is likely that man, who, probably because of his high degree of evolution, is more sensitive to the ill effects of radiation than lower organisms, will not come to harm.

The atmosphere can readily be contaminated by nuclear reactors. A pile, for instance, may be air cooled and the air passing through the reacting core will be exposed to bombardment by neutrons and made radioactive. Fortunately the main constituents of the air, oxygen and nitrogen, although they do capture neutrons, form radioactive isotopes which are small in amount and of very short half life. They do not, therefore, constitute a hazard. Argon is the element which is most seriously affected. The A^{40} is converted by capture of a neutron to A^{41} which emits both β- and γ-rays. This isotope has a half life of 1·8 hours. Because argon is a noble gas it does not enter into the composition of the body, but it can be taken up into the body fluids in solution. However, it can be calculated that this effect is less limiting than that exerted from the external irradiation of the body by the β- and γ-rays of the A^{41} in the atmosphere. A permissible concentration of A^{41} in the atmosphere can, therefore, be set by comparison with the so-called tolerance values for external γ-radiation.

Pollution of water supplies is similarly possible if the piles are water cooled. The cooling water, like the cooling air, in the example above, is bombarded with neutrons. The direct effect on the hydrogen and oxygen atoms again is negligible as far as the production of radioactive isotopes of these substances is concerned, but impurities in the water, such as sodium, may have a strong affinity for neutrons and become radioactive. This can be largely overcome by having purified water supplies and by delaying the effluent until short-lived radioactive isotopes have decayed.

As long as the fissionable material and its fission products are sealed in the pile there is no danger from them. Sooner or later, however, these have to be chemically processed to salvage the uranium, the plutonium and any other product of the fission that may be wanted. Some of the fission products are gases and others, notably I^{131}, are volatile. The escape of these gaseous products into the atmosphere and the discharge of the non-volatile unwanted fission products into the standard drains have to be prevented. The unwanted materials must be separated, concentrated and stored safely. No air filtration system is perfect, however, so some fraction of the volatile fission products must be expected to reach the atmosphere. Similarly the chemical apparatus must be washed and occasional spills must be anticipated : thus some fission products will reach the drains.

THE TOLERANCE DOSE

It is necessary, therefore, to consider what are the maximal quantities of these fission products that can be allowed to escape in this way without doing any significant damage to us. Given these figures the chemical engineers can design their plants accordingly. This consideration of tolerable doses of radioactive substances is obviously of importance to whole communities. The principles will, however, apply also to individuals who are the subjects of clinical investigation with radioactive isotopes.

From the medical point of view what one requires to know is that quantity of each radioisotope which can be held permanently by a person without producing any observable ill effect; that is, it shall not significantly affect his expectation of life or health, it shall not induce new growths and it shall not seriously damage his germ plasm.

The only certain way in which such a quantity could be determined would be by direct experiment which is obviously impossible. Therefore, it has at present to be assessed by indirect means. For this there are two possible methods. The one is the theoretical comparison of the probable effects of internally deposited radioactive substances with those of externally applied radiations. The other is a comparison

with the known effects of an internally deposited naturally radioactive substance : radium.

EXTERNAL X- AND γ-RADIATION AS A CRITERION OF TOLERANCE

This is the method which has hitherto been the one almost solely employed. The basis is the accumulated experience over the last 50 years with penetrating x-rays and the γ-rays of radium. The International Committee on x-ray and Radium Protection formerly made recommendations, after assessing the evidence, to the effect that a healthy person can tolerate with impunity exposure to one international roentgen per week. This was written into the recommendations[1] of the British x-ray and Radium Protection Committee, 1943. x-, γ- and similar radiations liberate electrons from the atoms in the media through which they pass. The resulting charged particles are, therefore, ions and the radiations which have this action are termed ionizing radiations. For x- and γ-rays the quantity or dose is measured by means of the resulting ionization. The unit of dose, the roentgen, has been defined as the quantity of x- and γ-radiation such that the associated corpuscular emission, per 0·001293 gm of air, produces in air ions carrying one electrostatic unit of quantity of electricity of either sign.

As a result of still more experience and from the evidence obtained from large scale experimentation with laboratory animals by the Manhattan District organization in the United States of America during the latter years of the last war it has been felt that this figure of 1r per week does not provide a sufficient safety factor. The Protection Sub-Committee of the Medical Research Council[2] in Britain has, therefore, recommended for the time being that ' in circumstances in which the whole body may be exposed, over an indefinite period, to x- or γ-radiation of quantum energy less than 3 Mev, the maximal permissible dose received by the surface of the body shall be 0·5 roentgen in any one week'. Opinion in the United States is substantially in agreement with this. There the maximal permissible weekly dose of 0·3r is to be measured not at the surface of the body but in free air and, therefore, without the contribution of scattered radiations.

The argument proceeds. The permissible dose of radiation for the whole body is limited by the organs or tissues which are most sensitive to damage by ionizing radiations. These are the haemopoietic tissues (the bone marrow and the lymphatic tissue), the gonads, the intestinal mucosa and the skin. In general, therefore, while the dose received by any organ should not exceed 0·5 or 0·3r per week, if an organ or tissue is known to be particularly resistant to damage by radiation,

an extra factor may be allowed for this when considering the concentration by such an organ of a radioactive isotope. Thus the thyroid gland is generally regarded as being resistant. It is known that the thyroid concentrates iodine when inorganic iodide is administered, so some would permit the thyroid tissue to receive from the fission product I^{131} a dose ten times that permitted to the bone marrow from, say, radioactive phosphorus.

The general principles of the theoretical considerations on determining the tolerance concentrations of radio-isotopes have been applied by COHN[3] to the calculation by formulae of the permissible concentrations of these substances in air, water or food. To quote Cohn,

> This relationship is expressed in terms of biological constants (degree of absorption, pattern of deposition, rate of excretion) and physical constants (half life, radiation energy) of the contaminating substance. Only β-radiation is considered, since γ-radiation, because of its low specific ionization is much less dependent upon localization for biological effectiveness. However, the formulation presented is easily modified to take both α- and γ-radiation into account.

The physical constants, even of the more or less recently discovered fission products, are known with considerable accuracy. SEABORG and PERLMAN[4] have reviewed them. The biological factors on the other hand are far from being constants in the best accepted version of the term. Within a given species of animal there is bound to be a considerable variation between individuals. Between species there is a notoriously wide variation.

Before the era of readily available radioactive isotopes, a good deal was known about the metabolism of most of the radicals found in the normal body and, from their therapeutic and toxicological properties, of many of the other elements. With the use of radioactive isotopes this knowledge has been extended, and, particularly due to the work of HAMILTON and the Berkeley school, much is known now about the metabolism of fission products. These are elements from the middle of the periodic table with mass numbers from 72 (Zn) to 158 (Gd). They were listed by the Plutonium Project[5] in 1946. There are two peaks in the graph of the fission yield with maxima about mass numbers 95 and 140 (*Figure 1*). The elements of reasonably long half life with appreciably high fission yields are given in *Table I.* Most biologists, after inspecting this list, would agree that there is a general ignorance of the biological fate of these elements. Therefore, to apply Cohn's calculations one is dependent on the excellent, but limited work of HAMILTON[6]

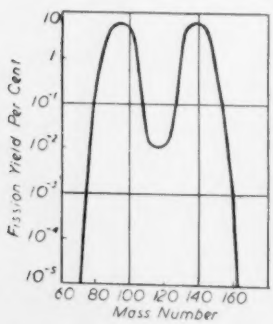


Figure 1. Yields of U^{235} fission products chains as a function of mass

Table I. Half lives of Elements with High Fission Yields

Isotope	*Half life*	*Isotope*	*Half life*
Kr^{85}	10 *years*	I^{131}	8 *days*
Sr^{89}	53 *days*	Xe^{133}	5·3 *days*
Sr^{90}	25 *years*	Cs^{135}	$2{\cdot}5 \times 10^4$ *years*
Y^{91}	57 *days*	Cs^{137}	33 *years*
Zr^{95}	65 *days*	Ba^{140}	12·8 *days*
Cb^{95}	35 *days*	Ce^{141}	28 *days*
Tc^{99}	4×10^5 *years*	Pr^{143}	13·8 *days*
Ru^{103}	42 *days*	Ce^{144}	275 *days*
Ru^{106}	1 *year*	Pm^{147}	3·7 *years*

and his colleagues with rats which has been published in the form of a summary and on other unpublished material.

One of Cohn's examples of the use of these general formulae is the calculation of the amount of a given product, inhaled or ingested daily from a source which is held constant (example a reservoir poisoned daily) which will give a maximum dose of 0·1 rep per day after infinite exposure. The term rep stands for roentgen equivalent physical. The roentgen, as has been noted previously, has been defined as a unit of dose for x- and γ-rays. For other ionizing radiations PARKER[7] has defined a unit of dose, the rep, as ' that dose of any ionizing radiation which produces energy absorption of 83 ergs/cc of tissue '.

Then
$$I_m = 60 . \frac{E}{W} \, q_0 ad . \frac{1}{\lambda + \beta}$$

where I_m = maximum dose rate of 0·1 rep/24 hours
E = average energy of β-particles in electron volts
W = weight of organ in grams absorbing all the β-radiation
q_0 = the unknown curies taken into the body per day
a = fraction of ingested or inhaled activity absorbed into the blood
d = fraction of the absorbed material being deposited in mass W of the given tissue
λ = decay constant of the β-emitter in days^{-1}
β = excretion constant of the β-emitter from the organ in days^{-1}

In a concrete example, the fission product Sr^{89}, the calculation would be

$$0{\cdot}1 = \frac{60(0{\cdot}5 \times 10^6) .}{10^4} \, q_0 . \frac{20}{100} . \frac{1}{(\cdot 0130 + \cdot 0015)}$$

as the average energy of the Sr^{89} β-ray is 500,000 ev, the Sr^{89} is deposited uniformly throughout the 10 kg skeleton, about 20 per cent of the ingested material is laid down in the skeleton, the physical half life is 53 days and the excretion constant is assumed to be ·0015 *i.e.*

$$q_0 = 2{\cdot}4 \times 10^{-6} \text{ curies}$$

Thus a daily intake of 2·4μc of Sr^{89} should, it is calculated, give a

local dose in the bone of 0·1 rep per day, which according to previously held ideas should be tolerable.

The main fallacy of these calculations is that, while absorption a and deposition d are variable with species and circumstances, it is possible to strike an average figure which would cover most circumstances. The insertion of a safety factor of 2 or 5 at the end of the calculation would ensure all cases being covered. But the factor β, the excretion constant, is a mathematical fiction. The excretion of all radicals is not random to give an exponential excretion. Excretion is rapid at first and slows down later but the ' excretion constant ' is at the best an approximation. For short lived isotopes the excretion constant can probably be neglected as it is likely to be small compared with the decay constant, but for the long lived isotopes *e.g.* Sr^{90}, half life 25 years, it will be the major component in the $\lambda + \beta$ factor and an error in the approximation will cause a significant error in the result of the calculation. The ultra conservative would therefore say, let us assume no excretion at all. This, however, may reduce the permissible amounts of the longer lived isotopes to prohibitively low values.

DEPOSITION OF RADIUM AS A CRITERION OF TOLERANCE

The other method of assessment of the quantity of radioactive isotope permitted in the human body is the possible comparison with radium. Just as there has been experience with the effects of external irradiation by x- and γ-rays, so also industrial hygienists have recognized ill effects from internally deposited radium. At the present time it is considered[8] from empirical evidence that 0·1μg permanently fixed in the body is the maximal permissible quantity. It is claimed by Evans *et alii*[9, 10] that burdens of the order of 1μg in the body have given rise to ill effects due to the radiations of radium. According to calculation this should be a reasonably safe amount. If one assumes that radium, like calcium, is concentrated in bone and, also like calcium, is uniformly distributed there, the local dose rate can be calculated. Each atom of radium on disintegrating to radon liberates an α-particle of very short range but of high energy (4·8 Mev). Each microcurie of radium ($\sim$ 1μg) by definition undergoes $3{\cdot}7 \times 10^4$ disintegrations per second. There are $8{\cdot}64 \times 10^4$ seconds in a day, $6{\cdot}25 \times 10^5$ electron volts per erg and 84 ergs/gm per rep. Therefore 1μg of radium distributed in a man's skeleton of 10 Kg weight should give a dose of

$$\frac{1}{10^4} \times \frac{3{\cdot}7 \times 10^4 \times 4{\cdot}8 \times 8{\cdot}64 \times 10^4 \times 7 \text{ rep/week}}{6{\cdot}25 \times 10^5 \times 84} = 0{\cdot}2 \text{ rep/week.}$$

Whereas 0·5 rep/week can be taken as the maximal permissible amount of x-, γ- and β-radiation there is general agreement that a safety factor of about ten should be applied when dealing with α-rays and their intensely local ionizing properties. For α-rays, therefore, the maximal permissible dose might be set at 0·05 rep, and the above calculation would suggest that the dose from the radium (1μg/per person) was 4 times greater than the permissible. On the other hand much of the energy of the disintegrations will be absorbed in the inorganic material of the skeleton and will not be liberated in the living cells of the bone. From the microscopic appearance of bone one would imagine this safety factor to be considerable. In making this calculation only two assumptions have been made

a that the radium is uniformly distributed in the bone

b that the dose to bone is the result of the disintegration only of the radium atom.

It will be remembered that radium is one of the elements in a radioactive chain. Each element produced by the disintegration of its parent element is itself radioactive. The chain is shown in *Figure 2*. Thus if the product of the disintegration of the radium atom *i.e.* radon, remained *in situ* until it also disintegrated it would liberate its energy to the same tissue cells as its parent, radium. The energy released in the tissue therefore would be not 4·8 Mev but 4·8 + 5·5 Mev. If all the atoms of the chain disintegrated *in situ* the whole of the energy of the chain would have to be taken into account. This would introduce an adverse factor of about 5. The balancing of these unknown factors might explain why 1μg radium distributed in the skeleton is dangerous.

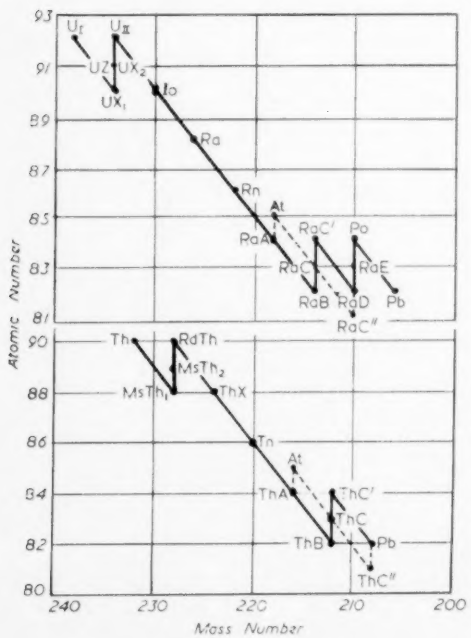


Figure 2. Decay chains of U^{238} *and* Th^{232}.

It is worth while, therefore, considering in some detail what is known about the metabolism of radium in the human body and the results of its retention.

METABOLISM OF RADIUM

There are a number of reviews in the literature on this subject, notably those by EVANS[9, 11]. Excellent though these are, it is difficult to get a clear cut picture of the problem. This is partly because a great deal of the evidence collected by Evans with collaborators is still unpublished. For the rest, much of the data concerns industrial intoxications and it is not always clear to what toxic substances the cases were exposed, in what doses and with what statistical results. In fact it is fairly typical clinical material about which as always it is difficult to be dogmatic.

Intake of radium into body— Radium has gained entrance to the body by ingestion, injection and inhalation. There was a time when radium was given internally by the former two methods as a therapeutic agent for such conditions as chronic arthritis, cancer, the reticulo-endothelioses, diabetes and a host of less explicable conditions. The vogue for this has fortunately passed. It has also been marketed as a ' quack ' tonic. Finally it has been ingested as luminous paint in the now notorious instances of the United States luminising industry of twenty to thirty years ago ; the operatives used brushes for applying the paint and pointed the brushes with their lips. Luminising is, of course, still carried on but brush work has been replaced with pen work and the hazards are now greatest from the inhalation of dust particles of the dried paint.

Complications arising in the determination of intake— Superficially, the hazards seem straightforward. However, whereas radium as a therapeutic agent seems in all cases to have been radium, the ' quack ' remedy, ' radithor ', which was marketed in the United States, was a mixture of radium and mesothorium (GETTLER and NORRIS[12]). Similarly, radioactive luminous paint as used in the United States seems originally to have been composed of zinc sulphide with traces of radium, but because of shortages and possibly other factors the radium was subsequently substituted partially or wholly with commercial mesothorium. CASTLE, DRINKER and DRINKER[13] certainly described the radioactive material as radium bromide, but HOFFMAN[14], reporting at the same time, talks of ' radium (mesothorium) necrosis '. These were the two earliest papers on the poisoning of dial painters. MARTLAND, CONLON and KNEF[15] shortly afterwards described the paint as radium-mesothorium paint, usually in the proportion of 20 : 80 and MARTLAND[16] makes it clear that the cases reported by Castle *et alii* were included in the larger survey of Hoffman. If one accepts that all these early cases of industrial poisoning by luminous compounds were due to a radium-mesothorium mixture, one has to conjecture whether the

radium or the mesothorium was the more hazardous. BLOOMFIELD and KNOWLES[17], writing in 1933, note that the mesothorium was not used in paint after 1930. Unless the commercial practice has changed since then, which is extremely likely, one may assume that industrial poisoning in dial painters commencing employment after 1930 is due to radium proper. It should be noted that radium and mesothorium are isotopes and so initially should be deposited in the body in the same fashion. Radium^{226} is a decay product of U^{238} with a half life of 1,600 years and emits an α-particle. Mesothorium I 228 is a decay product of thorium 232 with a half life of 6·7 years and emits a β-particle. The subsequent decay products are shown in *Figure 2*. It will be noted that radium decays first to the gas radon. This gas being readily soluble in body fluids can more or less easily diffuse to the circulating blood and so has a high probability of being excreted in the expired air. The body is then spared the energy of the subsequent decay products. EVANS[18] gives a mean figure of 45 per cent of the radon that is produced as being so excreted. HESS and MCNIFF[19] found a higher mean of 68 per cent and a greater scatter. Mesothorium I, on the other hand, decays to mesothorium II with a half life of 6·13 hours. This decays by β-particle emission to radiothorium, an isotope of thorium with a half life of 1·9 years. Radiothorium decays by α-particle emission to thorium X, another radium isotope, and this in turn decays to 54-second thoron, the isotope of radon. Because of its very short half life, a very much smaller proportion of thoron must be expired than is the case with radon, perhaps only a few per cent. Thus the maximal energy of decay per atom of radium in the body may be calculated as 4·8 (the energy of the radium α-particle) $+ \frac{55}{100} \times (24{\cdot}5 + 1{\cdot}7)$ (the energy of the α- and β-particles of the decay products) *i.e.* 19·2 Mev. This assumes that none of the decay products of radon is excreted. With mesothorium one would need to assume that practically the whole of the energy of the decay process, namely 33·3 Mev, was absorbed in the body. In both, of course, redistribution of the radioactive material may take place in the body because of the change of the atomic number. It would not be expected that a transmuted element would take the place of its parent element in the bone lattice. In the radium series the half lives of the decay products, radium A, radium B, radium C, radium C′ and radium C″ are of the order of minutes only or less. Radium D, an isotope of lead, however, has a long half life and so has radium F, polonium. MINOT[20], LOMHOLT[21] and BEHRENS[22] were early workers who had shown lead to be selectively taken up by bone. LACASSAGNE and LATTES[23], however, have shown that

polonium has no special affinity for bone and Lomholt reported similar findings for bismuth of which the isotope RaE has a half life of 5 days.

Short term retention of radium in the body— The radium burden of the body is calculated from two measurements

1 the gamma radiation from the body due to radium B and C, and

2 the analysis of radon in the breath.

The latter is obviously a measure of the radon loss from the body ; the former is a measure of the radon which does not escape but decays in the body.

The methods of EVANS[18] for the measurement of the γ-ray emission allow one to measure down to $0{\cdot}2\mu$g of radium in the body. HESS and MCNIFF[19] have claimed to have pushed the sensitivity to the uttermost limit to determine $0{\cdot}1\mu$g in the body. This γ-ray measurement does not discriminate between deposited radium and mesothorium. Analysis of the breath does permit this discrimination because of the difference in half lives of the two isotopes, radon and thoron, and in the case of dial painters both have frequently been found together (IVES *et alii*[25]).

Let us consider first of all what data are available on the cases who are known to have received radium. CAMERON and VIOL[26] in the early optimistic days of internal therapy with radium state ' even with the use of 1 mg doses we have, up to this time, never seen the slightest ill effect '. They note, however, that much of the radium given was retained in the body and quote figures of retention in various organs sampled at autopsy of a case of carcinoma of the uterus who had received 1 mg intravenously three months earlier. Bone marrow, liver and lungs were the organs which showed the greatest concentration per gram dry weight. Soon afterwards SEIL, VIOL and GORDON[27], did daily estimations of the radium excreted by a young man who had received 100μg intravenously. The rate of excretion fell from about 20 per cent per day to less than 1 per cent per day after one week. The excretion was noted to be mainly in the faeces and one half of the dose given had been eliminated at the end of the first week. Seil himself volunteered to take by mouth 50μg. The first experiment was spoilt by one of those unfortunate errors that results in faeces being thrown away, so he took another 50μg. After a week about 15 per cent of the ingested radium was still unexcreted and probably retained, as only small amounts were being discharged in the faeces at the time. We thus have a suggestive picture of strong retention by the body, 50 per cent and 15 per cent respectively after intravenous and oral administration. These figures are dependent on the stated doses having actually been given. In this connexion one must jump forward considerably to 1944 when EVANS and co-workers[10] carried out a

careful experiment with animals. They found that the doses actually administered were substantially lower than the projected because the radium tended to remain adsorbed on the walls of the ampoules and syringes, even in the presence of carrier. This type of radium dosimetry is thus very uncertain and the dose stated on the bottle is not by any means the dose given. The error of the dosage in Seil's experiments, if present, would show a greater retention than actually existed. SCHLUNDT and FAILLA[28] repeated Seil's volunteer act. Presumably, therefore, Seil had not in the 16 years' interval come to great harm. Schlundt, however, was more cautious : he drank water known to contain 2 to 5μg of radium and eliminated 91 per cent in 4 days. He repeated the performance two weeks later and 98 per cent was excreted in 5 days. This almost complete elimination, 91 and 98 per cent respectively, is impressive, but the figures of the amount taken, namely 2 to 5μg, are rather vague. However, this is probably loose reporting rather than loose measurement !

Long term retention of radium in the body— On a longer term basis what proportion of the radium originally taken remains permanently fixed, and when it is permanently fixed, at what rate is it eliminated ? SCHLUNDT and FAILLA[28] give some data on the latter point. True, their two reported cases were radium dial painters, but for one it is specifically stated that only radium and not mesothorium was involved. The woman had a measured body content of 24μg of radium and a daily elimination of 1,600$\mu\mu$g, which gives a daily coefficient of elimination of 0·005 per cent. Most later authors agree on the order of this coefficient.

On the former point of the long term retention of radium, EVANS *et alii*[10] make a note that in people taking a radium nostrum, 0·1 to 10 per cent of the amount ingested is fixed permanently in the body. They promise further details by AUB and EVANS, but so far they do not appear to have been published. Obviously such fundamentally important studies on man are urgently required. Some data are also available from BARKER and SCHLUNDT[29]. They collected ten cases who had in the past been treated with radium chloride. The total doses given and the time which had elapsed since the last dose varied widely. One man, for instance, had taken 1,455μg orally over two and a quarter years and yet one month after the last dose only retained about 4 to 8μg. Two cases on the other hand, who had received 200 and 300μg approximately up till two years previously, showed no measurable retention, while two cases, who had received 900 and 1,100μg intravenously up till twelve months or so previously, still retained a minimum of 19 and 10μg respectively. Retention is thus rather varied, but allowing for the errors of dosage mentioned and

the relative insensitivity compared with later methods of the measurements of retained radium, it is obvious that radium can be extensively eliminated. This is confirmed by a further investigation by SCHLUNDT, NERANCY and MORRIS[30]. They treated mentally deranged patients with periodic intravenous injections of 10μg doses of radium for 2 to 6 months, giving total doses of 70 to 450μg. Estimations of radium contents of their bodies were made at the end of the course and six months later. An average of 4·3 per cent of the projected dose was measured as retained at the end of the course and 1·9 per cent six months later. These results are in the expected range, between those of Seil *et alii* for the early days after a single administration and those of Schlundt and Failla with long-standing radium retention. The qualitative picture emerges that much of the radium is immediately excreted. Only a small proportion, or in fortunate cases negligible quantities, are firmly retained with an extremely slow turnover.

Mechanism of retention of radium in the body— The behaviour of radium is probably due to its location within the body. Radium is first, as we have seen from CAMERON and VIOL[26], distributed not only in bone but fairly widely in the soft tissues, especially the liver and lung. At this stage the radon emanation appears to be able to get fairly readily into the blood stream. SCHLUNDT *et alii*[30] were impressed that at the time of their first measurement at the completion of the course of radium only one seventh of the radon and its decay products was held in the body. This is to be compared with the average of 55 per cent retention quoted by EVANS[18] in cases of chronic radium retention (presumably largely dial painters) and the approximately 50 per cent retention reported by IVES *et alii*[25] in dial painters, and of BARKER and SCHLUNDT[29] in radium treated patients. Thus even though in the early stages the bulk of the radium is already located in bone it must be in bone well supplied by the circulating blood, that is the trabecular bone. Later, it becomes increasingly buried in ivory bone or in bone inaccessible to the circulation.

ELIMINATION OF RADIUM

Granted that radium is ultimately mainly deposited in the bone, 99 per cent or so, as one would expect from its position in the periodic table and relationship to calcium, is it possible to speed the natural elimination? FLINN and SEIDLIN[31], impressed with recently discovered properties of parathyroid hormone, gave to three cases of chronic 'radium poisoning', with jaw lesions, a month's course of parathormone in doses up to 40 units on alternative days. After two courses, they claimed a reduction of the radium content of three dial painters to 50 per cent. Later FLINN[32] reported equally striking results from the

administration of viosterol (vitamin D_2). Both these publications, as preliminary communications, give only scanty details of the cases. As no later detailed reports have been traced from these authors, their results must be regarded as suggestive only. CRAVER and SCHLUNDT[33], moreover, could not confirm the results. AUB and EVANS and their colleagues[34] give a most convincing story, though on three cases only, but the calcium and radium balances were obviously carried out in an expert fashion. In two of these three cases with a long-standing radium burden, only one being a dial painter, low calcium diets, alone and with supplemental thyroid, ammonium chloride and parathormone treatment, did increase the excretion of radium several-fold, but did little to reduce the absolute total radium in the body. The doses of parathormone used were vastly greater than those of Flinn and Seidlin. Irradiated ergosterol produced no increase in the elimination of radium. In the third case who had recently acquired his radium, the dietary and therapeutic regimen seemed to have a beneficial effect on the excretion, possibly because of the relative accessibility of the radium.

There is an interesting calculation in this paper[34]. For one of the cases of chronic radium retention there were figures for the concentration of radium in the blood and for the daily output of radium in the urine. From these figures one can calculate that the plasma clearance is of the order of 0·1 ml per minute, and that only 0·01 per cent of the circulating radium is cleared by the kidney each day. This must mean that radium is either strongly reabsorbed by the tubules from the glomerular filtrate, or, alternatively, that it is bound in the plasma and is not filtered off by the glomeruli. The hypothesis of a binding of the radium in the plasma is the more attractive one, as it might account also for the failure of the gut to excrete the radium. These figures throw a somewhat new light on the problem of the decontamination of the body. There is not only a difficulty in mobilizing the radium from the bone, but a still greater difficulty in excreting it when it is mobilized. As this difficulty is not nearly so pronounced in the acute stage of radium intoxication, it certainly seems that in the chronic state the radium is in some bound form.

The conception of a recycling of the radium fits in with certain known facts on the unevenness of distribution of radium as observed autoradiographically (MARTLAND[16] and EVANS *et alii*[10]). Part of the non-uniformity may well have been due to the fact that the radium originally deposited in areas where calcification was exceptionally active remained there. But Evans *et alii* have noted that when fractures occurred long after the radium had been fixed, radium

concentrated at the site of healing. The radium must thus have been available in the fixed deposits, even though it was not being excreted.

EFFECTS OF RETENTION OF RADIUM

Having discussed the absorption, retention and excretion of radium from the body, let us consider the results of radium retention. As far as possible, the author has attempted so far to keep to a discussion of radium alone and not of radium-mesothorium mixtures of luminous paint. At this point one must revert to the luminisers and the physicists and chemists who have been exposed to mixed sources and who have subsequently developed clinical abnormalities. The clinical cases described in the literature where one can be certain that radium only has been administered are too few, but it seems clear that these cases do not differ qualitatively from the far larger group of luminisers.

The first report of intoxication amongst luminisers appears to have been noted by a dentist, BLUM[35], calling attention, during a discussion on osteomyelitis of the mandible and maxilla, to the condition of 'radium jaw'. Bone necrosis of the jaw with infected sinuses has been extensively described subsequently. While the jaws are the bones most noticeably affected, other bones show clinical and radiological abnormalities. The flat bones of the skull and shoulder girdle and the spine and weight-bearing bones seem particularly to be involved. The skiagrams show areas both of abnormal translucency and abnormal density. In experimental animals to whom radium has been administered, all the long bones showed areas of increased density adjacent to the joints with thinning and loss of density of the shafts of the bones (THOMAS and BRUNER[36]; EVANS *et alii*[10]). Spontaneous fractures, both in the experimental animals and in the luminisers, have been common. Histologically, bone necrosis is seen, that is loss of osteoblasts from the bone lacunae. Osteoclasts are also destroyed so the necrotic bone sometimes remains unabsorbed. DUNLAP *et alii*[37] noted the deposition of fibrillary bone on the surfaces of the unabsorbed necrotic bone.

This so-called osteitis or bone necrosis seems to occur in those more acute types of radium poisoning. Coincidentally or even in advance of bone symptoms there is an anaemia. MARTLAND, CONLON and KNEF[15] noted it in luminisers and REITTER and MARTLAND[38] in a radio-chemist. While the peripheral blood suggested an aplastic anaemia such as had been observed by MOTTRAM[39] and WEIL and LACASSAGNE[40] the marrow was in fact hyperplastic. The photographs and photomicrographs of MARTLAND[16] are incontestable. In experimental animals, both aplasia and hyperplasia of the marrow have been found, sometimes both together (THOMAS and BRUNER[36]), aplasia

in the ribs, vertebrae *etc*, and hyperplasia in the shafts of the long bones.

In the more chronic type of case, there is the added complication of osteosarcoma. This was first described by MARTLAND and HUMPHRIES[41]. In this and others of Martland's cases, chemical analysis demonstrated that most of the deposited radioactive material was mesothorium. A later review by MARTLAND[42] revealed more cases of osteosarcoma, but in these only radium and not mesothorium was detected. Osteosarcoma has been produced in rats by EVANS *et alii*[10] and reported in more detail by DUNLAP *et alii*[37]. This was an experiment with radium. SABIN, DOAN and FORKNER[43] had earlier produced experimental osteosarcoma, not only with radium but also with mesothorium. They suggest that mesothorium was somewhat the more active, but it is not clear whether purified mesothorium or commercial mesothorium containing radium was used.

To sum up we have the following information :

a There is no collected series of cases where one can be sure that the ill effects in all are due to radium alone and not to some other radioactive compound.

b Even if one neglects the luminisers, one cannot be certain in the human cases or in experimental animals precisely what doses of radium have actually been given.

c One may accept as an approximation that of radium taken into the body about 1 per cent is retained in the bones.

d This radium in bone is fixed and excretion of 50 per cent from the body takes about 40 years.

e The radium though fixed in the bone may be more mobile than is generally accepted.

f The complicated decay system renders the understanding of radium metabolism difficult and militates against precise calculation.

g Experimental animals appear to be much less sensitive to radium than man.

The following doses were given in various experimental series :

SABIN, DOAN and FORKNER[43] : approximately monthly intravenous injections of 5μg of $RaCl_2$ into rabbits. The animals received, therefore, about 70μg before death occurred after about 1½ years. One animal had a sarcoma.

THOMAS and BRUNER[36] : subcutaneous injections of 5μg of $RaCl^2$ at irregular intervals to rats with a total dose of 40 to 60μg in 117 to 191 days. Deaths occurred at 261 to 338 days.

ROSENTHAL and GRACE[44] : for 3 months daily oral $RaSO_4$ suspension given to rabbits : a total of approximately 100μg. Deaths occurred within 3 to 18 months.

Of these Thomas and Bruner were the only ones who make any mention of the retention of the radium given. 15 to 30 per cent was measured as retained 9 to 69 days after the completion of the course.

Evans, Harris and Bunker[10] and Dunlap *et alii*[37] executed more exact experiments and in the latter publication, when the object of the experiment was induction of sarcoma with small doses, they record that rats fed 100μg of $RaCl_2$ retained an average of 2μg at the end of 10 months. Nine of their 13 animals developed malignant neoplasms. Thus while none of the experiments has been planned to reduce the dose to a tolerable amount *Table II* gives derived concentrations which have been shown to be toxic for experimental animals. It is notable that these concentrations are far higher than those which have produced damage in man.

Table II. Toxic Concentrations of Radium

Authors	*Possible* Ra *content of animal* (μc)	*Possible conc. of activity/gram of bone* (μc/gm)
Sabin *et alii*	5–10	$2–4 \times 10^{-2}$
Thomas and Bruner	10	4×10^{-1}
Rosenthal and Grace	1	4×10^{-3}
Dunlap *et alii*	2	6×10^{-2}

Evans *et alii*[10] from the results of their investigations put the rat/man sensitivity ratio for radium as 1/150. They also note that the rat expires an average of 85 per cent of its radon compared with the 45 per cent average in man.

It is notable that with this guesswork the activity per gram of bone is by far the lowest in the experiments of Rosenthal and Grace. These workers administered $RaSO_4$, a substance many times more insoluble than $BaSO_4$. It is surprising, therefore, that the rabbits absorbed any at all. The symptoms suggest an acute intoxication. It may be that as much as 10 or 20 times the suggested amount was in fact absorbed. The whole is a dramatic manifestation of the potency of radioactivity. Whereas barium, a notoriously toxic metallic radical, exerts no action even when enormous doses of $BaSO_4$ are given by mouth, radium as $RaSO_4$ was rapid in producing its ill effect.

In spite of these disadvantages radium will still remain one of the radio-isotopes of choice for animal experimental investigation. The γ-rays emitted by its decay products allow some assessment of the location of the radium to be made externally. The fact that it decays primarily to radon, much of which is expired, allows one to measure small quantities held by the body.

Above all for the present purpose, of the naturally occurring radioactive substances, infinitely more is known about its radiotoxic action in man than that of the others. Uranium is known to be chemically toxic but there are no reports of damage from its radioactive properties.

There is one case report by MacMahon, Murphy and Bates[45] of a sarcoma produced by thorium and several papers on the production of neoplasms in experimental animals (Roussey and Guerin[46] and Selbie[47]).

RADIUM COMPARED WITH FISSION PRODUCTS

For radioactive fission products that might be expected to metabolise in a fashion similar to radium *e.g.* Sr^{89}, Sr^{90}, Ba^{140}, one can make more or less direct comparisons with radium. If the permissible amount of radium in the human skeleton is 10^{-1} μg, this is equivalent to approximately 10^{-5} μc/gm. The energy absorption per gram of bone is thus 5×10^{-11} c Mev if one considers only the energy of the disintegration of the radium atom or 15×10^{-11} c Mev if one includes a proportion of the energy of the disintegration of the decay products. In comparing other radioactive substances with radium it is wise only to make assumptions that are pessimistic and for this calculation one will neglect both the possible contribution of the decay products of radium and the relative biological efficiency of the α-particles compared with β-rays. Thus let us say the maximal permissible amount of energy to be released per gram of bone is 5×10^{-11} c Mev. When we come to Sr^{89} which has an average energy of 0·5 Mev for its β-particles, 5×10^{-11} c Mev is released per gram by 10^{-4} μc and in a 10 kgm skeleton this would represent a total burden in the skeleton of 1 μc. It is notable that, in spite of the pessimistic assumptions this calculated Ra/Sr activity ratio of 1/10 for equal effects, is the same as that found experimentally by Brues, Lisco and Finkel[48] in mice. For Sr^{90} and Ba^{140} the calculation is complicated by the fact that, unlike Sr^{89} which decays to a stable atom, these isotopes have daughter products Y^{90}, La^{140} which are radioactive. Once again to be pessimistic one must assume the worst, and here this means including the energy of the disintegration of the daughter substance with that of the parent. Thus one would take the average energy of Sr^{90} and Y^{90} to be $\frac{1}{3}$ (0·6 + 2·3) = 1 Mev and that of Ba^{140} and La^{140} to be 0·9 Mev. The calculated permissible burdens of either of these fission products would thus be about one half that of Sr^{89}. There is no published experimental work as with Sr^{89} to show how valid are these calculations for Sr^{90} or Ba^{140}.

Whereas Ba and Sr will be expected to be deposited diffusely in bone like radium and calcium, there are other fission products which, though they deposit in bone, do so not uniformly as do radium and strontium but in concentrated form on the periosteal and endosteal surfaces. No quantitative assessment has yet been made of the mass of volume of bone in which these materials are laid down. By

inspection of autoradiograph patterns of the deposition one may assess this volume pessimistically at 1/10 of the whole skeleton or 1 kg instead of 10 kg. HAMILTON[6] has shown that yttrium and cerium deposit in this way. Taking, therefore, the tolerable concentration of Y^{91} (β-rays, average energy 0·5 Mev) in bone equal to that for radium *i.e.* giving 5×10^{-5} μc Mev/gm, this activity is 10^{-4} μc/gm or 1/10 μc per effective skeleton of 1 kg. The uncertainty of this assessment lies not only in the factor for unevenness of distribution, but more particularly in the metabolic fate of the yttrium. Radium, we have seen, though fixed seems capable of movement from one area of bone to another. Yttrium like plutonium is according to Hamilton associated with the protein matrix of the bone and may be truly fixed in position. This may be a safety factor. As COPP *et alii*[49] have pointed out radioactive material originally deposited subendosteally becomes buried by further laying down of bone. Thus the very sensitive bone marrow cells are increasingly screened. On the other hand individual osteoblasts are presumably involved by this activity for the duration of their life and may be stimulated to undergo neoplastic change.

Finally there are fission products which do not localize in bone *e.g.* Xe^{133}, Cs^{137}, Ru^{103} and Ru^{106}. There is little justification for comparing these with radium. The permissible concentrations of these substances are, therefore, best calculated from the amount of energy they contribute to the body as a whole referred to the so-called maximal permissible dose of 0·5r per week for irradiation of the whole body.

SUMMARY

The maximal permissible or tolerable amount of any radioactive isotope within the body can be calculated roughly.

The basis for the calculation may be either the tolerance dose of external x- or γ-radiation or the tolerable burden of radium. It is suggested that the latter has some advantages. The literature on radium metabolism and intoxication is therefore once again reviewed.

REFERENCES

[1] *Sixth Revised Report* British x-ray and Radium Protection Committee (1943)
[2] *Seventh Revised Report ibid* (1948)
[3] COHN, W. E. *Nucleonics* 3 (1948) 21
[4] SEABORG, G. T. and PERLMAN, I. *Rev. Mod. Phys* 20 (1948) 585
[5] *Plutonium Project J. Amer. chem. Soc.* 68 (1946) 2411
[6] HAMILTON, J. G. *Radiol.* 49 (1947) 325
[7] PARKER, H. M. *Advances in Biological and Medical Physics* 1 New York, 1948 p 243
[8] *U.S. National Bureau of Standards Handbook H27* (1941)
[9] EVANS, R. D. *J. Ind. Hyg.* 25 (1943) 23
[10] — HARRIS, R. S. and BUNKER, J. W. M. *Amer. J. Roentgenol* 52 (1944) 353
[11] — *Amer J. Publ. Hlth* 23 (1933) 1017

REFERENCES

[12] GETTLER, A. O. and NORRIS, C. *J. Amer. med. Assn* 100 (1933) 400
[13] CASTLE, W. B., DRINKER, K. R. and DRINKER, C. K. *J. Ind. Hyg.* 7 (1925) 371
[14] HOFFMAN, F. L. *J. Amer. med. Assn* 85 (1925) 961
[15] MARTLAND, H. S., CONLON, P. and KNEF, J. P. *ibid* 85 (1925) 1769
[16] — *Arch. Path. and lab. Med.* 2 (1926) 465
[17] BLOOMFIELD, J. J. and KNOWLES, F. L. *J. Ind. Hyg.* 15 (1933) 368
[18] EVANS, R. D. *Amer. J. Roentgenol* 37 (1937) 368
[19] HESS, V. F. and MCNIFF, W. T. *ibid* 57 (1941) 91
[20] MINOT, A. S. *J. Ind. Hyg.* 6 (1924) 125
[21] LOMHOLT, S. *J. pharm. and exp. Ther.* 40 (1930) 235
[22] BEHRENS, B. *Deut. Med. Woch.* 60 (1934) 890
[23] LACASSAGNE, A. and LATTES, J. S. *C.R. Acad. Sci., Paris* 178 (1924) 488
[24] LOMHOLT, S. in *Radioactive Indicators* London, 1948
[25] IVES, J. E., KNOWLES, F. L. and BRITTEN, R. H. *J. Ind. Hyg.* 15 (1933) 433
[26] CAMERON, W. H. and VIOL, C. H. *Radium* 4 (1914–15) 57
[27] SEIL, H. A., VIOL, C. H. and GORDON, M. A. *ibid* 5 (1915) 40
[28] SCHLUNDT, H. and FAILLA, G. *Amer. J. Roentgenol* 26 (1931) 265
[29] BARKER, H. H. and SCHLUNDT, H. *ibid* 24 (1930) 418
[30] SCHLUNDT, H., NERANCY, J. T. and MORRIS, J. P. *ibid* 30 (1933) 515
[31] FLINN, F. B. and SEIDLIN, S. M. *Johns Hopk. Hosp. Bull.* 45 (1929) 269
[32] — *J. Amer. med. Assn* 96 (1931) 1763
[33] CRAVER, L. F. and SCHLUNDT, H. *ibid* 105 (1935) 959
[34] AUB, J. C., EVANS, R. D., GALLAGHER, D. M. and TIBBETS, D. M. *Ann. intern. Med.* 11 (1938) 1443
[35] BLUM, T. *J. Amer. Dent. Assn* 11 (1924) 802
[36] THOMAS, H. E. and BRUNER, F. H. *Amer. J. Roentgenol* 29 (1933) 641
[37] DUNLAP, C. E., AUB, J. C., EVANS, R. D. and HARRIS, R. S. *Amer. J. Pathol.* 20 (1944) 1
[38] REITTER, G. S. and MARTLAND, H. S. *Amer. J. Roentgenol* 16 (1926) 161
[39] MOTTRAM, J. C. *Arch. Radiol. Electrother.* 25 (1920–21) 194
[40] WEIL, P. E. and LACASSAGNE, A. *Bull. Acad. méd. Paris* 93 (1925) 237
[41] MARTLAND, H. S. and HUMPHRIES, R. E. *Arch. Pathol.* 7 (1929) 406
[42] — *Amer. J. Cancer* 15 (1931) 2435
[43] SABIN, F. R., DOAN, C. A. and FORKNER, C. E. *J. exp. Med.* 56 (1932) 267
[44] ROSENTHAL, M. and GRACE, E. J. *Amer. J. med. Sci.* 191 (1936) 607
[45] MACMAHON, H. E., MURPHY, A. S. and BATES, M. I. *Amer. J. Pathol.* 23 (1947) 585
[46] ROUSSEY, G. and GUERIN, M. *Pr. méd.* 49 (1941) 761
[47] SELBIE, F. R. *Brit. J. Exp. Path.* 19 (1938) 100
[48] BRUES, A. M., LISCO, H. and FINKEL, M. MDDC 145
[49] COPP, D. H., AXELROD, D. J. and HAMILTON, J. G. *Amer. J. Roentgenol* 58 (1947) 10

9

MECHANICAL PROPERTIES OF FIBRES AND MUSCLES

M. G. M. Pryor

THERMODYNAMICS OF ELASTIC STRAINS

A MUSCLE is an engine for the conversion of chemical energy into mechanical work, operating by means of reversible strains in the working substance, the muscle proteins. This kind of power cycle has not been much investigated, and before considering muscle in particular, the general principles involved will be discussed at some length with the object of discovering how many different kinds of cycle are possible and how they may be distinguished.

If a body is subjected to a small elastic strain at constant temperature, the work done upon it, dW, may be analysed into two components ; that associated with changes in internal (potential) energy, dU, and that associated with changes in heat content, dQ. Then

$$dW = dU + dQ \qquad \ldots.(1)$$

where dQ is the amount of heat given up by the body to its surroundings. If the change in heat content is very small, it is clear that most of the work done is stored as potential energy ; the restoring force must therefore be due mainly to attractive forces between the atoms. If there are large heat changes associated with strain *i.e.* if dQ is large, the part played by potential energy is correspondingly reduced ; it follows that the restoring force must be largely derived from the thermal (kinetic) energy of the atoms. If strained at constant temperature, a body of this kind can only store energy as a decrease in entropy.

Characteristics of normal and rubbery solids— Using this criterion, it is possible to classify all solids into two kinds according to their behaviour under reversible strains ; those in which the temperature is very little affected by rapid deformations, and those in which it is considerably increased. The first class, which will be called normal, includes metals and most other solids. Typical examples have a small reversible extensibility of the order of 0·01 per cent and a high Young's modulus of the order of 10^{12} dynes/sq cm. The second class, which will be called rubbery solids, includes rubber and some other high polymers. They usually have a large reversible extensibility of the order of 100 per cent, and a low Young's modulus of the order of 10^{7} dynes/sq cm.

Special thermodynamics of rubbery elasticity— The first exact treatment of the thermodynamics of rubbery elasticity is due to WIEGAND and SNYDER[1], since when the subject has been much developed by MEYER and his school (see reviews by MEYER[2], and WÖHLISCH[3]). From the first and second laws, an expression is derived which makes it possible to deduce the changes in internal energy that occur when a rubbery material is stretched from observations of the thermal coefficient of tension at constant length. For small isothermal strains in an ideal rubber

$$K = (\partial U/\partial l)_T + T(\partial K/\partial T)_l \qquad \ldots.(2)$$

where K is the applied load, l the length and T the absolute temperature. Within a small range of temperature K can be expressed as a linear function of T *i.e.* the curve can be replaced by its tangent

$$K = a + bT \qquad \ldots.(3)$$

where $a = (\partial U/\partial l)_T$ and $b = (\partial K/\partial T)_l$.

The value of $(\partial K/\partial T)_l$ is determined by measuring at various temperatures the tension in a sample held at a fixed length ; time must be allowed for the tension to reach a steady value, and an allowance has to be made for normal isotropic thermal expansion. When $(\partial K/\partial T)_l$ is known, the value of $(\partial U/\partial l)_T$ is obtained from equation 3. If within a given range K varies in proportion to the absolute temperature, then $K = bT$, and $a = (\partial U/\partial l)_T = 0$; there is no change in internal energy on straining, and all the work done on the body is transformed into heat. If $K > bT$, $(\partial U/\partial l)_T$ is positive; there is an increase in internal energy on straining, and the heat liberated is less than the work done. Finally when $K < bT$, so that $(\partial U/\partial l)_T$ is negative, there is a decrease in internal energy on straining, and the heat liberated is greater than the work done. This rather paradoxical condition is found in raw or lightly vulcanized rubber. The internal energy decreases because the molecules pack closer together when they are aligned parallel by the applied stress ; the extra heat liberated is therefore analogous to the latent heat of crystallization of a normal solid. This interpretation is confirmed by observations of the x-ray diffraction pattern, which shows a progressive change from an amorphous to a typical fibre diagram as the material is extended.

Calculations of entropy— All rubbery solids consist of long flexible chain molecules, and many attempts have been made to find an exact relation between elastic properties and molecular weight by calculating the entropy associated with different configurations of the chains. To make the calculation easier, it is assumed that the material consists of a loose three dimensional network of chain molecules. The chains

are supposed to be flexible, as a consequence of free rotation between atoms, and it is also assumed that they do not attract one another or interfere with each other's motion. For such an ideal rubber it is possible to calculate the probability that the two ends of a chain containing any given number of links will be separated by any given distance. If we take the ends of the chain to represent points of intersection in the network, then there will be a linear relation between the distance separating the ends of a chain and the dimensions of the material in bulk. Hence it is possible to obtain a relation between the strain in the material and the probability of the configuration of the chains associated with such a strain. The probability is related to the entropy by Boltzmann's constant, and thus the entropy associated with a given strain can be found ; from this the change in entropy, and hence the work, required to pass from one state to another can be calculated. If changes in cross-sectional area associated with longitudinal strain are neglected, the modulus of elasticity can be calculated in terms of the molecular weight, the absolute temperature, and various constants characteristic of the material. The results of several such calculations agree approximately with experiment, although they differ from one another ; a review is given by MEYER and VAN DER WYK[4].

Calculations of entropy are not very useful in considering the mechanism of muscular contraction because they can give no information about the intermediate states, in which we are mainly interested. Without some mental picture of the molecular mechanism concerned it is very difficult to imagine the working of a power cycle based on rubbery strains or to predict the effects of departures from ideal conditions such as the existence of strong intermolecular attractions.

Statistical-mechanical treatment—A useful picture of the molecular mechanism responsible for the restoring force in strained rubber can be built up by statistical arguments, starting from a consideration of the forces acting on a single chain. If we consider an ideal flexible chain molecule subject to random impacts from all directions, it appears that all parts of the chain remote from the ends will be under tension. In a simple system such as rubber the kinetic energy of the chains is derived mainly from the impacts of neighbouring chains ; in a strongly solvated system the impacts of solvent molecules may also be an important source of energy. The maximum load in any segment of chain will depend on the number of contiguous atoms which move in the same direction at any one time ; if we assume that the probability of two adjacent atoms moving in the same direction is small, then the probability of any considerable number moving in the same

direction is very small indeed. From this it follows that the maximum tension will be reached for a comparatively short chain and will not be affected by increases in length above this limit. In a large population, we may consider every chain as subjected to a mean tension f. The value of f is not substantially affected by the position or configuration of the chain, and is no greater when the ends are far apart than when they are close together, a necessary condition if the internal energy is to remain unaffected by strain in the material. Müller[5] has published a similar statistical treatment, but makes the assumption that tension in the chains depends on their configuration, which is incompatible with the conditions for an ideal rubber.

If a chain makes an angle θ with the normal to any plane passing through the material it will contribute at that plane a resultant tension T in the direction of the normal equal to $f \cos \theta$.

When the material is not strained, we may suppose that this tension is balanced by an equal and opposite bombardment pressure exerted on any surface in contact with the material ; if the chains are sufficiently mobile to exert a tension in the direction of their axes, they must be capable of delivering impacts in directions at right angles. If the mean pressure exerted by a chain is q, the resultant pressure p in the direction normal to the plane will be equal to $q \sin \theta$.

Angular distribution of links— To extend this treatment to a large aggregate of chains we must consider their angular distribution. To represent the orientation of a link on a polar diagram two properties must be known, direction and sense. Direction defines the angular coordinates of a link without making any distinction between an angle of θ and one of $\theta + 180°$; sense distinguishes between these two alternatives. Sense can only be defined by an arbitrary convention. In the following discussion it will be represented as measured from the lower end of any chain toward the upper end ; values of θ are measured clockwise, starting from the top.

In the unstrained material there will be an equal number of links lying in every direction. When the material is strained by a force acting in the direction $\theta = 0$, the angular distribution will be distorted but will remain radially symmetrical about $\theta = 0$. If the chains continue in the strained condition to change direction under the influence of random impacts, but are connected to each other at widely separated points, the angular distribution of their links at any moment will exhibit a random scatter about the preferred direction, and may be represented by the normal statistical relation, the standard deviation decreasing as the material is extended.

The angular distribution is conveniently represented by a three dimensional polar diagram, in which the length of the radius r is

proportional to the number of links lying in that direction. For a random distribution

$$r = \frac{1}{\sqrt{2\pi}.\sigma} \exp\left(-\theta^2/2\sigma^2\right) \qquad \ldots\ldots(4)$$

where σ is the standard deviation. Instead of the Gaussian formula it is more convenient to use the relation

$$r = \frac{1}{\sqrt{2\pi}.0{\cdot}36b} \cos^8\left(\theta/b\right) \qquad \ldots\ldots(5)$$

where b is a constant whose value decreases with increasing strain. If r is plotted against θ this gives a curve sufficiently similar for our purposes to the normal curve, with a point of inflection corresponding to the standard deviation at $\theta = 0{\cdot}36b$. If the material is compressed, instead of being extended, in the direction $\theta = 0$, or if it is subjected to a uniform dilation in the plane at right angles to $\theta = 0$, as in the skin of an inflated spherical balloon, the distribution of the chains will be normal about a plane at right angles to $\theta = 0$. This distribution can conveniently be represented by substituting $\sin^8(\theta/b)$ for $\cos^8(\theta/b)$ in expression 5 above.

The shape of the curves obtained by plotting r against θ for different values of σ are shown in *Figure 1*. When σ is large the curve goes more than once round the origin. Values of θ greater than $\pi/2$ can be interpreted by considering the configuration of individual chains. When σ is less than 0·36 radians *i.e.* when the curve reaches the centre at a value of $\theta < \pi/2$, no part of any chain is recurrent ; when σ is less than 0·72 ($b < 2$) there may be recurrent portions, but no loops, and when σ is greater than 1·44 ($b > 4$) loops may be present. The chains are only capable of a certain minimum radius of curvature, so that the greatest number of successive loops that can occur in any chain *i.e.* the maximum possible value of σ, is limited by

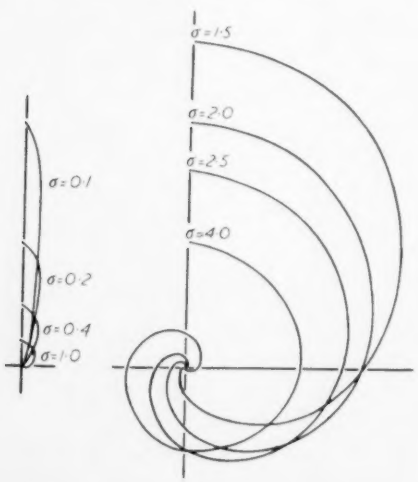


Figure 1. Angular distribution curves : normal statistical distribution curves (equation 4) plotted on polar coordinates for different values of standard deviation σ. Using polar coordinates $\int_0^\infty r\, d\theta$ *is constant, so that the area enclosed becomes smaller as the value of σ increases. To show their shape more clearly, the curves for higher values of σ have been plotted on a larger scale*

the chain length. As the direction only, and not the sense, is important in calculating the load contributed by any chain, the sign of cos θ in the equation $T = f \cos \theta$ may be neglected.

The curve obtained by plotting r against θ represents a section through the axis of the three dimensional distribution figure; if the radii are evenly spaced, the number generated by rotating any radius about the axis $\theta = 0$ must be proportional to the circumference of the figure at the level at which the radius in question cuts it *i.e.* to $2\pi(r \sin \theta)$. Hence the total number of links making an angle θ with the direction of tension is proportional to

$$n = r(r \sin \theta) \qquad \ldots.(6)$$

The chance of any link cutting a plane whose normal is $\theta = 0$ depends on the length of its projection in the direction $\theta = 0$, hence the number cutting such a plane is proportional to $n \cos \theta$.

Relation of load to randomness— Combining these expressions we can now obtain an expression for the total tension acting in the direction $\theta = 0$. Where b is less than unity ($\sigma < 0{\cdot}36$ radians), the total tension is proportional to

$$\mathcal{Z} = f \int_0^{2\pi} r^2 \cos^2 \theta \sin \theta d\theta \qquad \ldots.(7)$$

By a similar argument, the pressure P is proportional to

$$P = q \int_0^{2\pi} r^2 \cos \theta \sin^2 \theta d\theta \qquad \ldots.(8)$$

The net tensile force is

$$K = \mathcal{Z} - P \qquad \ldots.(9)$$

When b is less than unity the length of a chain is equal to the sum of the projected lengths of all its links *i.e.* to

$$L = l \int_0^{2\pi} r^2 \cos \theta \sin \theta d\theta \qquad \ldots.(10)$$

where l is the length of a link (see *Figure 2*).

If we assume a limited number of chains of fixed length we can neglect changes in cross-sectional area provided that b is less than unity. If it is required to express the tension as a stress, the breadth B of the sample is proportional to

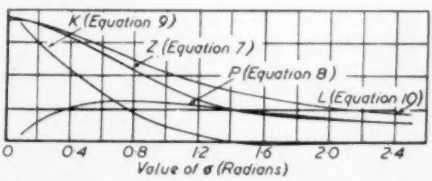


Figure 2. Values of $\mathcal{Z}$, P, L and K from equations 7 to 10 plotted against the standard deviation of the angular distribution of the links in the high polymer chains

$$B = \int_0^{2\pi} r^2 \sin^2 \theta d\theta \qquad \ldots.(11)$$

When b is greater than unity, and some of the chains are folded back on themselves or looped, the number of links crossing any plane is increased because some chains cross more than once. For any given value of b the length will be less and the cross-sectional area greater than that calculated from equations 10 and 11. The tension K will be increased by a factor proportional to the increase in cross section ; this factor cannot easily be calculated from the distribution curve.

The information available from equations 8 to 11 can be summarized graphically by plotting K against L, taking arbitrary values of f and q such that K tends to zero as σ increases. The curve shows that dK/dL increases at high strains, as found in experiments on rubber (*Figure 3*). The value of σ gives a measure of the randomness of the chains. The variation of K or L with σ is interesting because of the importance which has been given to the idea of a tendency to assume a more random distribution in calculations based on consideration of entropy states. At low values of σ, K decreases rapidly as σ increases, but at values of σ in the neighbourhood of 1 radian ($b = 1/0{\cdot}36$) the curve flattens out, and for high values of σ the value of K is practically constant (*Figure 2*). A value of σ little greater than 1 radian corresponds to almost complete randomness of angular distribution if sense is neglected ; very much larger values are necessary to bring about randomness of both sense and direction. A perfectly random arrangement of chains according to both sense and direction ($\sigma = \infty$) could in fact only be attained for infinitely long chains.

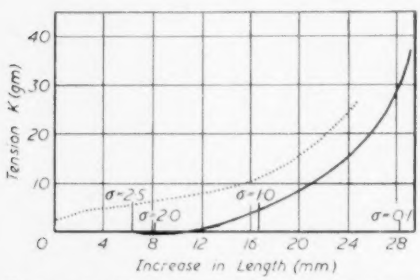


Figure 3. Load/length curves. The upper dotted curve is for a strip of lightly vulcanized rubber; the lower continuous curve was obtained by plotting K against L (equation 9)

Modulus of elasticity and chain length— All calculations based on statistical estimation of entropy lead to expressions in which Young's modulus appears as a function of chain length. MEYER and VAN DER WYK[4] have objected to this as being contrary to common experience, and have produced figures to show that for many materials there is no simple relation between molecular weight and elastic properties. This may be true, and still not prove that the calculations are unsound in principle. Many of their examples are materials which at room temperature show properties intermediate between normal and rubbery solids (see p 225), so that their elastic moduli

as measured at any appreciable rate of strain contain a component due to internal energy changes, and will not correspond to the calculated values, which refer only to the rubbery component. Leaving this difficulty out of account, and confining our attention to the rubbery component of the observed tensions, the statistical-mechanical approach suggests that although the extent of the greatest reversible strain is a function of chain length, the tension exerted at the maximum extension is not, so that both $d\sigma/dL$ and dK/dL increase with decreasing chain length. Stated in this form the dependence of elastic modulus on chain length seems less objectionable. The way in which the elastic properties of rubber vary with the degree of vulcanization agrees with this interpretation. In this instance the effective chain length is the mean distance between cross links ; as the chain length is decreased by increasing the proportion of cross links, extensibility is diminished and the value of dK/dL (Young's modulus) increased. At a sulphur content of 1 per cent, Young's modulus is about 14×10^6 dynes/sq cm (reversible extensibility of up to six times) ; at 40 per cent sulphur content Young's modulus increases to about 5×10^{10} dynes/sq cm and the maximum reversible extensibility decreases to less than 1 per cent.

ASTBURY[6] has stated that the maximum extension possible by straightening out a random tangle of links, each of length l, is $l \cos 45°$, but this is only true of single links not connected together into chains.

Effect of viscosity— If whole chains are free to slip past one another, no configuration other than a completely random one can be stable at any strain. In practice however, for materials of high molecular weight this limitation is not important, because relative movement of whole chains is opposed by a considerable viscous resistance even when the small movements of parts of chains involved in changes of configuration are comparatively free. Why this should be is clear from simple statistical considerations. The magnitude of the energy barrier opposing movement between two sections of chain in contact depends on the number of contacts between consecutive pair of atoms, and this in turn is a function of the mean value of r (equation 5) and will increase as the orientation becomes more perfect. At large strains therefore, when relative movement between whole chains might be expected to be most rapid, the viscous resistance is at its highest. If the molecular weight is low, the viscous resistance never becomes very high, and rubbery properties are not conspicuous, because the orientation on which rubbery stress depends decays rapidly.

It is convenient to distinguish between the viscous resistance opposing relative movement of parts of chains, and that opposing the movement of whole chains, although at large strains the two merge

into one another ; the former will be called internal and the latter true viscosity.

Imperfect rubbers and viscous-elastic effects— To extend these calculations to imperfect rubbers it is necessary to consider how the existence of cohesive forces between chains will affect the mean tension f in the chains and the mean pressure q they exert. We may consider each chain as surrounded by a hollow cylinder of neighbours. Every movement of the atoms at right angles to the long axis of the chain is subjected to a retarding force, which decreases the tension produced in the chain in much the same way as intermolecular attractions in an imperfect gas decrease the force of impacts on the wall of the containing vessel. The decrease in the value of f and q may be represented by a constant m analogous to van der Waals's constant a. The value of f increases with temperature ; if m remains constant the value of $(f - m)$ will decrease with decreasing temperature until it reaches zero at the ' rubbery freezing point', below which the material behaves as a normal solid.

It is characteristic of materials in which m is large that they exhibit pronounced elastic after effect : they resist any sudden change of shape by considerable forces but will respond to a small sustained load by large deformations ; when load is removed they return slowly to their original shape. As an example we may consider the behaviour of a strip of polyvinyl chloride in tension, shown in *Figure 4*. The specimen is held at a fixed length until the tension is constant, and is then allowed to shorten rapidly for a small distance A and kept at the new length. The tension suffers an immediate large decrease, but recovers gradually to nearly its original value. Comparable results are obtained by a sudden small stretch B ; the tension at first increases very much but later drops.

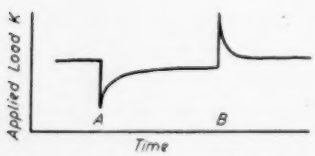


Figure 4. Viscous-elastic behaviour of a typical slow rubber

The mechanical properties of a material like this may be represented by a model consisting of springs and dashpots, arranged either in series or in parallel, or both. The behaviour of most materials can be represented by a sufficiently complicated combination of springs and dashpots, although for some it becomes necessary to introduce modifications by which the stiffness of the springs and the viscosity of the dashpots can be made to increase with increasing strain. These models facilitate the description of a material, but they offer no explanation of the events they imitate, and provide no analogy for several important properties of the real material, as for example the correlation of internal viscosity with elastic modulus, or the dependence of both on

temperature. In spite of these objections, many people have tried to account for elastic after effect by supposing that materials which exhibit it do actually contain structures corresponding to springs and dashpots. MEYER and FERRI[7] for example have suggested that rubber consists of two phases, a flexible sponge responsible for the tension, and a viscous fluid that flows through its meshes.

On the statistical-mechanical theory the occurrence of elastic after effect can be explained without supposing the existence of two phases, if we assume that where two chains touch they form weak secondary valence bonds. These bonds cannot resist any considerable force for any considerable time, but over a short period they can sustain quite large forces. When the material is suddenly stretched it behaves as if it were intensely cross linked ; the effective chain length is small, so that a small increase in length suffices to cause a considerable improvement in orientation (see p 223), and hence a considerable increase in tension. If the load is sustained, relative movement between chains occurs, and they take up the equilibrium configuration proper to their true chain length; the value of b rises and the tension falls. For a quick release the reverse is true.

The extreme cross linked condition, whether temporary or permanent, may be considered as one in which relative translational movement between parts of chains is restricted, and elongation of the material occurs principally by rotation of links about their mid points. The effect of a rapid change of shape on the angular distribution diagram may therefore be represented by a conformable transformation, in which circles become ellipses by an extension of the figure in one direction; extension being in the direction $\theta = 0$ for a quick stretch, and in a direction at right angles for a quick release. When b is less than unity the figure remains approximately normal, although the standard deviation departs from the equilibrium value for that length. If the curve goes more than once round the origin it ceases to be normal when distorted, and is transformed by a stretch into the shape of a cigar, and by a quick release into that of a lens. The distribution obtained by a quick stretch will give a larger value for K than the equilibrium distribution, and that obtained by a release a smaller value ; in either case the tension will slowly return to its equilibrium value as the angular distribution returns to normal.

Behaviour of frozen rubbers— Rubbery solids at temperatures near their ' rubbery freezing point ' may be considered as extreme types of imperfect rubbers, in which the value of $(f - m)$ is nearly zero. Their behaviour is interesting for two reasons, first because it reproduces at slow rates of loading the behaviour to be expected of typical rubbery solids at very high rates of loading, and secondly because it makes it

possible to bring into a common scheme the peculiar elastic properties of materials like hair, in which great extensibility is combined with a potential energy mechanism (Woods[8, 9]).

As an extreme type we may consider *Perspex* (polymethylmethacrylate). At room temperature *Perspex* behaves like a normal solid, with a Young's modulus of about 3×10^{11} dynes/sq cm, and maximum reversible strain of the order of 0·001 per cent. If a tensile specimen is carefully polished so that there are no surface cracks at which failure might originate, it shows a fairly well defined yield point, above which it necks and extends like a ductile metal. At room temperature this extension is not regained when load is removed, but if the specimen is immersed in water at about 80°C it rapidly returns to its original shape, showing that the ' plastic ' extension was in fact rubbery. The applied stress has been sufficient to overcome the energy barrier opposing relative movement between parts of chains, and they have been pulled out to a more extended configuration. At room temperature they remain in the extended configuration when the load is removed, because the restoring force is zero, but at temperatures above the 'rubbery freezing point' they return to their original random arrangement. Although the force opposing extension is derived entirely from internal energy, the restoring force responsible for complete return to the original length is derived from entropy; the potential energy mechanism will only cause return to the nearest potential trough, a matter of a few thousandths per cent strain. For any finite rate of strain at room temperature, the modulus of elasticity will appear to be high, as in normal solids.

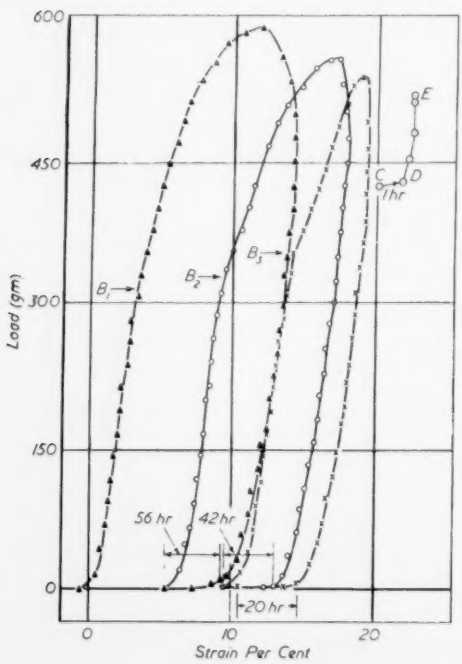


Figure 5. Three successive cycles of loading and unloading on the same filament of polyvinyl chloride, diameter 0·3 mm. Readings at intervals of 1 min except where marked. B_1, B_2, *and* B_3 *are successive yield points.*
After the third cycle, the filament was loaded rapidly (30 sec) to C, and then allowed to creep to D, the load being maintained approximately constant; after 1 hr it was further loaded to E

Sample kindly supplied by Messrs. I.C.I. Plastics.

Polyvinyl chloride behaves in a similar way. In *Figure 5* is shown a load/length curve for a filament of polyvinyl chloride strained fairly slowly. There is a well-defined yield point B above which extension is obviously dependent on time and so appears to be plastic. The restoring force at room temperature is just measurable however, and when the load is released the material very slowly returns to its original length.

It is not necessary to suppose that more than a very small fraction of the load on the upward limb of the curve is due to kinetic effects. As the value of m is high, the rubbery tension in the chains $(f - m)$ will be very small. If the chains stick together where they touch, a very slight increase in the length of the material as a whole is sufficient to align them nearly parallel, although their distribution may not be normal (see p 225). Further elongation of the material is then resisted by elastic forces derived from strain energy. The additional tension will impose shear on the bonds between chains, which they can only sustain up to a certain point, above which they yield and allow the material to extend until the chains are entirely unravelled. Once the chains are extended more or less parallel, elongation is resisted by a strain energy mechanism again, and the modulus again increases. This second increase in the elastic modulus can be made to appear at loads only slightly above the yield point if the material is held at that load until extension is complete, as shown by DE in *Figure 5*. This type of load/length curve has been described in several kinds of synthetic high polymer by LEADERMAN[10].

The load/length curve for hair (*Figure 6*) is like that for polyvinyl chloride. There is a false yield point, the elastic modulus increases again at high extensions, and there is the same slow return at zero load. With mechanical properties so much alike it seems reasonable to suppose that the molecular mechanism involved in the two materials is similar. ASTBURY and WOODS[11] explained the shape of the curve as due to the successive extension of three different kinds of

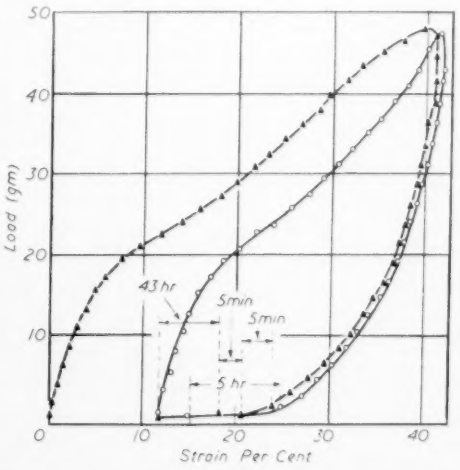


Figure 6. Two successive cycles of loading and unloading for a human hair, diameter 0·1 mm. All readings at intervals of 1 min except where marked

keratin arranged in series; the three kinds being intracellular keratin, enclosed within elongate cells; keratin forming the external walls of the cells, and intercellular keratin cementing the cells together. In a later paper[12], Woods rejects this explanation, and supposes instead that the cells themselves contain all three phases, one as a continuous network and the other two in its meshes. It is doubtful whether such an arrangement would in fact produce the mechanical effects observed, but in any case it seems to be an unnecessarily complicated hypothesis; if polyvinyl chloride, which is amorphous and monophasic, can yield the same type of curve, it is simpler to suppose that the curve for hair is due to the intrinsic properties of a single material.

As in polyvinyl chloride, the slow return of hair to its original length at zero load is accelerated by a rise in temperature. Agents which decrease molecular cohesion, such as heating in steam, treatment with alkalis *etc* lower the yield point, reduce the value of the elastic modulus on the upward arm of the curve, and accelerate return at zero load. Esterification of polar side chains has the same effect (Harris, Mizell and Fourt[13]). Meyer and Haselbach[14] find that for elephant's hair swollen in thioglycollic acid $(\partial K/\partial T)_l$ is positive. This seems to be in agreement with what is known of the effect of other swelling agents, and does not conflict, as they claim, with the conclusion of Woods, that under normal conditions most of the restoring force is derived from potential energy.

Astbury and Bell[15] interpret the crystal structure of α-keratin as an arrangement which gives the closest packing of side chains. If we consider the angular distribution of links corresponding to Astbury and Bell's model of α-keratin it appears that there are equal numbers at $\theta = 0$ and at $\theta = 90°$, but none at angles greater than $\theta = 90°$, a condition not far removed from complete uniformity for direction without uniformity of sense. The α-crystalline arrangement may thus be considered as a modification, induced by attractive forces between side chains, of an intermediate condition in the curling up of the chains. It is metastable, and if sufficient heat is supplied to melt the crystallites the angular distribution develops further towards the equilibrium condition in which there is uniformity of both sense and direction; the additional curling up causing supercontraction of the material in bulk. Astbury and Bell observe that by distorting valency angles the side chains could be packed a little more closely still, but this itself will not account for a supercontraction to two thirds of the original length. Astbury[6] has pointed out that the large changes in length observed in supercontracted hair and in muscle could be accounted for if the folds in the molecular ribbons present in the α-configuration were extended laterally, as in the crossed β-configura-

tion described by RUDALL[16] for epidermal proteins. In the crossed β-configuration there are more links at angles of $\theta = 90°$ or 180° than at $\theta = 0$, so that the angular distribution figure approximates to that for a rubber compressed in the direction $\theta = 0$. Such an arrangement might arise from a more isotropic distribution by a process of crystallization, but in that event it seems unlikely that it could be concerned in muscular contraction, because crystallization is necessarily too slow. MERCER[17] has recently found evidence for the existence of the crossed β-configuration in wool fibres swollen by a mixture of sodium sulphide and urea solution.

Energy relations of frozen rubbers— The mechanical ' rubbery yield point ' of these ' frozen rubbers ' has a thermal analogue, the ' rubbery melting point '. Below the rubbery melting point tension at constant length decreases slightly with increasing temperature because of the normal isotropic thermal expansion of the material ; above it tension increases with temperature. The relationship between tensile load and temperature is most easily understood by considering a comparable gaseous system ; a liquid near its boiling point enclosed in a rigid vessel. As the temperature is raised the pressure will increase in the manner shown in *Figure 7*, which is compiled from the figures given for the saturation vapour pressure of water in Landolt-Bornstein's tables. If excess liquid were always present the pressure should rise as shown in curve AB, but if the volume of water originally present is limited, at some stage in the process it will have evaporated completely, and from then on the pressure will be a linear function of the temperature, BC.

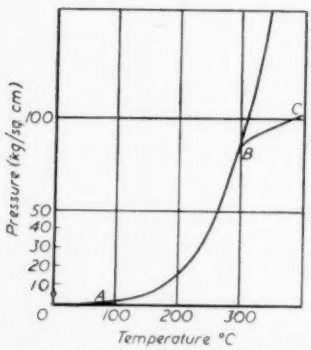


Figure 7. Saturated vapour pressure of water at various temperatures

Changes in the elastic properties over a range of temperatures about the rubbery melting point have been investigated in several materials and in all of them the relation between isometric tension and temperature is similar. *Figure 8* shows two curves for tensile load at constant length plotted against temperature for a nylon filament in water. Similar curves have been obtained by WÖHLISCH and DU MESNIL DE ROCHEMENT[18] for tendon, for elastoidin by PICKEN[19] and for gutta percha by MEYER[20].

WÖHLISCH and DU MESNIL DE ROCHEMENT have applied to such curves a modified form of the Clausius-Clapeyron equation

$$(\partial T_m/\partial K)_l = T_m(\partial L/\partial \phi_m)_T$$

where T_m is the temperature of the rubbery melting point, Q_m the latent heat of the process of melting, and K the stress. If the material shortens when heated, the right hand side of the equation must be positive, and

$$(\partial T_m/\partial K)_l > 0$$

i.e. the rubbery melting point must rise with increasing load.

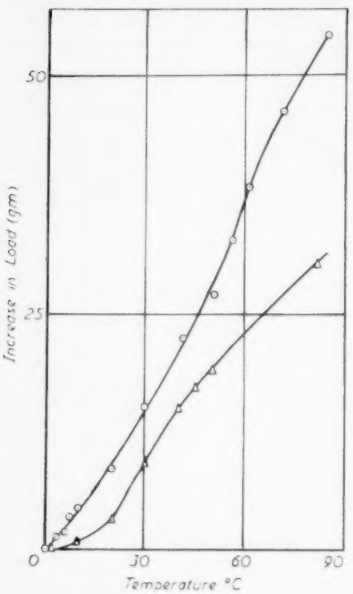


Figure 8. Increase in load exerted at constant length by nylon filament in water at different temperatures. Upper curve, starting from zero load; lower curve, starting at a load of 170 gm (c 25 kg/sq cm)

It is interesting to note that the rubbery melting point does not exactly correspond either to the boiling point of a liquid or to the melting point of a crystalline solid. The occurrence of a crystalline phase does not materially alter the shape of the curve; if values for ice are substituted for those for supercooled water in *Figure 7* the curve is comparatively little changed. The parameter which corresponds in an imperfect rubber to the volume of liquid originally present is presumably a factor related to the mean number of atoms which cooperate to produce the mean tension f.

WOODS[8, 9] has investigated the energy relations of hair, nylon, and casein wool under load at different temperatures, but without taking account of the existence of the rubbery melting point. He determines the value of $(\partial K/\partial T)_l$ from measurements of load at two temperatures only (20° and 40°C) and then calculates the contribution K_s of entropy effects to the total load K at 20° by using the relation

$$K_s = T(\partial K/\partial T)_l$$

This implies that $(\partial K/\partial T)_l$ is constant, and as it is not, his conclusions are sometimes misleading. If the rubbery melting point of the material he is investigating lies above 40° his values may be approximately correct, but when it lies within the range 20° to 40°, as it does for nylon and casein wool, he obtains too high a value for $(\partial K/\partial T)_l$ and in consequence too high a value for K_s. His conclusion that nylon and casein wool belong to a peculiar class of 'hyperentropic' materials need not therefore be accepted.

Perspex and other synthetic plastics also provide an analogy to the

'set' of hair. If *Perspex* is strained at a temperature above its rubbery melting point (*ca* 70°C) and is cooled in the strained position, it will retain the new shape indefinitely, provided that it is not again heated above the rubbery melting point. This is analogous to 'temporary set'; 'permanent set', which persists even at temperatures above the rubbery melting point, can be imposed only by heating the material above its true melting point (*ca* 150°C). In hair the position is complicated by chemical factors; the effect of raising the temperature may be reinforced by chemical changes which weaken the lateral bonds between protein chains. SPEAKMAN[21] has shown that permanent set is favoured by conditions which lead to the breakdown of the cystine links originally present and their re-establishment in the new configuration. In terms of our previous argument (p 224) this is to decrease m instead of increasing f.

The similarity in the shape of the load/extension curves for hair, which is partially crystalline at all extensions, and polyvinyl chloride, which is almost entirely amorphous, suggests that the occurrence of crystalline phases at any stage in the extension of a material is not by itself of great importance for the mechanical properties. When intermolecular cohesive forces are strong, the chains will tend, at any length of the material, to settle down into the configuration which permits of closest packing. This may be crystalline, but the occurrence of one or more crystalline phases during the extension of a material cannot be taken as evidence that the restoring force is derived from internal energy.

POSSIBLE ELASTIC POWER CYCLES

We are now in a position to consider the workings of elastic power cycles in more detail. From what has been said of the nature of elastic strains, it appears that only two fundamentally different kinds of cycle are possible, one involving variations of internal energy and the other variation of entropy: we shall consider them in that order. The distinction between the two kinds of cycle has been clearly expressed by RAMSEY[22].

Potential energy cycles— A potential energy engine could be constructed with any long chain polymer in which there were mutually attractive groups disposed at intervals along the molecules, the intensity of the attractive force (and hence the elastic modulus of the material) being variable. It is easy to imagine ways in which the attraction between permanent dipoles might be varied chemically, either by affecting the magnitude of their charge or the dielectric constant of the intervening medium. When the attractive forces are strong, successive centres will approach one another and the chain

will fold together and shorten. To complete the cycle the attractive forces might be decreased and the chains pulled straight again. The cycle can be represented on an indicator diagram whose axes represent load K and length L (*Figure 9*).

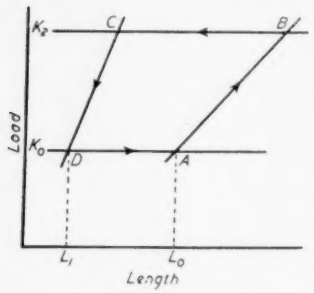


Figure 9. Indicator diagram for a potential energy cycle

Starting at A under a load K_0, the load is increased to K_2 and the material extends to B. The material is then allowed to shorten, and the elastic modulus is increased progressively to maintain the load constant until a new length L_1 is reached at C. The modulus of elasticity is now kept constant, and the load reduced to K_0 again, allowing the material to shorten to D. The material is then extended to its original length L_0 at A, progressively reducing the elastic modulus so as to keep the load constant at K_0. The work done per cycle is represented by the area ABCD.

A mechanism of this kind has been implied by several people and has been discussed in detail by MEYER[23, 24]. WEBER[25] has pointed out one difficulty, but there are in fact two serious objections. The first concerns the magnitude of the strains involved. Attractive forces between point charges fall off as the inverse square, or higher power, of the distance between their centres, and it is difficult to imagine how such forces can act except over very small distances. Yet a smooth muscle can shorten to about a third of its resting length. To account for such large strains we must either suppose the existence of some new kind of long range attraction, which is possible but unlikely, or provide a mechanical system to magnify the motions of the contractile substance. SZENT GYÖRGYI[26] has suggested an arrangement which would have the required effect. He supposes that the micelles of the contractile protein, myosin, are each closely attached throughout their length to ribbon-shaped aggregates of a non-contractile protein, actin. When the myosin contracts, the composite strip, actomyosin, will magnify the movement in the same way as a bimetallic strip magnifies the thermal expansion of a metal. It is doubtful whether actin and myosin do combine in quite this way however (see PERRY and REED[27]), and the necessity for something of the kind remains as an unwelcome complication to all hypotheses of the potential type.

The second difficulty concerns the direction in which the forces act. If an increase in attractive forces is to cause longitudinal contraction, successive centres along the same chain must approach one another

without too much lateral approximation of neighbouring chains. Intermolecular forces are not directional, so that to ensure this effect a very regular structure would be required ; such a structure is possible, but the optical properties and x-ray diffraction pattern of contracted muscle do not in fact suggest great regularity.

RISEMAN and KIRKWOOD[28] propose a mixed type of cycle. They suggest that the protein chains are held extended in the resting muscle by forces of repulsion between charged centres spaced along them, and curl up under the influence of thermal agitation when these forces are reduced. This is open to objection from both sides ; there seems to be no reason why the repulsive forces should not cause equal expansion in all directions, and it leads one to expect that heat will be absorbed during contraction.

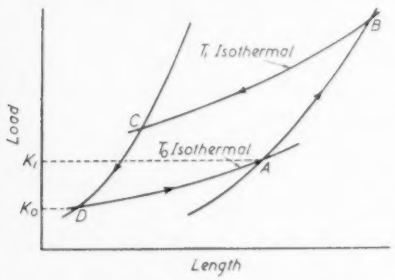


Figure 10. Indicator for a kinetic energy (entropy) cycle

Entropy cycles: heat engines— An engine operated by thermal (kinetic) energy can be constructed from any material in which tension at constant length is increased by a rise in temperature. A simple heat engine can be made by suspending a weight from a strip of lightly vulcanized rubber hanging between two electric radiators. If the load and the intensity of the radiation can be controlled, and the rubber is fitted with a removable insulating jacket, this system can be taken round a reversible cycle analogous to Carnot's cycle (*Figure 10*).

Starting with the rubber at a temperature T_0, under a load K_1 at A, the insulating jacket is put on and the load increased adiabatically until the temperature rises to T_1 at B. The insulating jacket is then removed and the radiators switched on ; the load is decreased until the rubber has shortened to a point C, sufficient heat being supplied to keep the temperature constant at T_1. At C the radiators are switched off, and the insulating jacket replaced, the load is decreased to K_0, and the rubber allowed to shorten adiabatically until the temperature has fallen to T_0 again. The insulating jacket is then removed, and the load gradually increased to its original value K_1, the operation being carried out so slowly that the heat produced is radiated away and the temperature remains constant at T_0. This cycle, a modification of the self-energizing pendulum described by WIEGAND[29] in 1925, can be performed on any material in which the tension/length adiabatics are steeper than the isothermals, that is, in

which the thermal coefficient of tension at constant length $(\partial K/\partial T)_l$ is positive.

The amount of work done per cycle depends, other things being equal, on the heat capacity of the working substance. If the rubbery melting point of the working substance lies between the working temperatures, or if it crystallizes when stretched, the work done per cycle will be increased. The effect is analogous to that produced in the Carnot cycle by using a saturated vapour instead of an ideal gas as working substance. Neither of these modifications affects the efficiency of the process, as can be seen from the entropy/temperature diagram.

Crystallization makes possible a special case, in which $K_1=K_0=0$; the working substance elongates spontaneously without the application of any external force. The place of the external force is taken by attractive forces between chains, and the elongation is accompanied by a fall in internal energy.

Chemical engines— For an imperfect rubber the tension at any length may be increased as well by a decrease in the attractive forces between molecules as by a rise in temperature. The strength of intermolecular cohesion can be controlled chemically by the addition of plasticizers. The commonest plasticizers used commercially to soften synthetic high polymers are complex esters, dibutyl phthalate, glyceryl sebacates *etc* which become adsorbed by their polar groups on to charged groups in the polymer, thereby concentrating the field of force of the charge within a smaller radius, and also by their bulk preventing a close approach of the chains. Their effect may be to some extent specific to particular substrates which the spacing of their polar groups happens to fit. For strongly solvated polymers, as for example proteins swollen in water, the amount of bound solvent will also affect the mean tension in the chains, an increase in the bound solvent tending both to increase the thermal energy of the chains and to decrease their cohesion. Ions which affect the swelling of a fibre may therefore be expected to affect its elastic properties as well ; this is probably the explanation of the contraction observed by MONGAR and WASSERMAN[30] in alginate fibres when calcium ions were replaced by sodium.

By varying the concentration of plasticizer or by changing the species of ion present, a simple chemical engine could be made, whose working was analogous to that of the heat engine described above. Plasticizer is added or removed where heat is added or removed in the heat engine; along the adiabatics no plasticizer is allowed to enter or escape from the system, and the isothermals become plasticizer isobars along which the vapour pressure or concentration of the plasticizer is kept constant. An engine of this kind can be constructed with any material in which

the tension at constant length is increased by the addition of plasticizer *i.e.* as long as $(\partial K/\partial[P])_l$ is positive. The efficiency can be assessed from the analogue of the temperature/entropy diagram ; for temperature take plasticizer vapour pressure or concentration $[P]$, and for entropy the quantity of adsorbed plasticizer divided by the concentration $(P/[P])$. Thus it appears that the efficiency should depend on the difference between the two working concentrations divided by the lower $([P]_1 - [P]_0)/[P]_0$. The consequences of crystallization of the working substance are the same as for a heat engine.

MODEL OF CHEMICAL ENGINE

The detailed working of an engine based on rubbery strains is complicated, and for simplicity's sake the properties of an inanimate model will be described in detail before the results are applied to the investigation of muscle.

To distinguish in practice between potential and entropy cycles an apparatus has been used in which the specimen is subjected to a tension of rhythmically varying intensity and the resulting deformations recorded. If the frequency is varied over a wide range, both the elastic modulus and the internal and true viscosities can be deduced from the same set of observations. The theory of this kind of measurement has been studied by physicists interested in the mechanical properties of rubbery solids *e.g.* TUCKETT[31]. There is also a review by ALFREY and DOTY[32]. The same method, or modifications of it, has been applied to muscle by WINTON[33] and by Buchtal and his school.

As the response of rubbery solids to applied loads is not instantaneous, it follows that if the material is subjected to a fluctuating load, the amplitude of the resulting deformation will decrease with increasing frequency. The apparent modulus of elasticity as deduced from the amplitude of the response in such experiments has been called the dynamic modulus. At low frequencies it approximates to the modulus as measured in static experiments, and at high frequencies it increases to a maximum which approaches the range of values typical of normal solids. The relation between frequency and dynamic modulus is not linear, the increase of dynamic modulus with increasing frequency being confined to a range of frequencies which will be called the critical range.

A fall in temperature or an increase in the intensity of intermolecular cohesion will tend to shift the critical range towards lower frequencies, because both changes tend to increase the internal viscosity. At frequencies below the critical range *i.e.* when the material is behaving wholly as a rubbery solid, either of these influences will tend to decrease the dynamic modulus, because of their effect in decreasing

the static (rubbery) modulus. At frequencies above the critical range, when the material is behaving as a normal solid, their effect will be negligible. It ought therefore to be possible to find out whether intermolecular forces have increased or decreased by comparing the relation between frequency and dynamic modulus at two stages in a cycle. Suppose the curve *R* in *Figure 11* applies to a material before it contracts; if contraction is due to an increase in intermolecular forces, the critical frequency should be reduced as in curve *I*, but if it is due to a decrease, the critical frequency should be increased as in curve *II*.

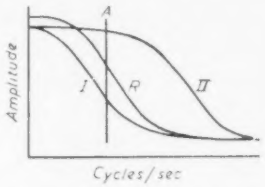


Figure 11. Variation of dynamic modulus with frequency

At frequencies within the critical range, say at *A* in *Figure 11*, the dynamic modulus should be increased on contraction in a potential and decreased in a kinetic cycle. In the experiments on synthetic plastics reviewed by Alfrey and Doty, the test specimens were alternately stretched and compressed by equal amounts, so that there was no resultant mean stress. In this way irreversible viscous flow is avoided, and the interpretation of the results in terms of springs and dashpots is made easier. For fibres or muscles the apparatus has to be modified in such a way that the specimen is always under tension. Under these conditions the specimen may undergo irreversible extension during the course of an experiment; if it does, the periodic deformations due to the oscillating load are superimposed on a shifting base line whose slope is a measure of the rate of viscous flow. When the viscous strain becomes large the specimen hangs slack between its attachments and no longer follows exactly the stresses imposed on it, but in the initial stages the slope of the base line provides interesting additional information. True viscosity is affected by the same factors as the internal viscosity, and we should therefore expect the slope of the base line to be steepest when the internal viscosity is lowest. In a

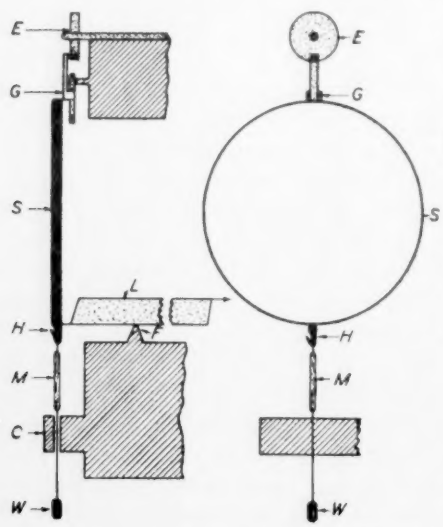


Figure 12. Apparatus for measuring the dynamic modulus

potential cycle the slope should decrease during the active phase, and in a kinetic cycle it should increase.

The apparatus used in the author's experiments, shown in *Figure 12*, consisted of a ring spring *S* with a high natural frequency, suspended vertically from an eccentric gear *E* driven by a motor through a set of gears which could be adjusted to give a range of frequencies from 5 cycles per second to 15 minutes per cycle. The spring was kept in place above by a guide ring *G* running on a vertical rod, and below by a hinged lever *L* which permitted only movement in the vertical plane. The lever hinged about a point *F* 15·5 cm behind the lower edge of the spring, and extended back behind this for a distance of 36 cm, where it ended in a point writing on a smoked drum. The movement of the lower edge of the spring was thus recorded with a magnification of × 2·3. For frequencies above 2 cycle/sec the part of the lever behind the hinge *F* was removed and the movement of the spring observed by means of a beam of light reflected from a mirror attached to the end of the lever immediately above *F*. On the lower edge of the spring was a hook *H* for the attachment of the specimen.

The lower end of the specimen was held in a clamp *C*. When it was required to immerse the specimen in test solutions, a bath was made by threading the specimen through the bottom of an inverted rubber teat.

To investigate the behaviour of the model under nearly isometric conditions, an apparatus was used which recorded small changes in length magnified 26 times on a smoked drum (*Figure 13*). The recording lever *L* is a light pith rod, hung on a thread from a firm wooden stand, and balanced about its point of suspension by a lead weight *W* at one end. The thread was attached above to a cotton-reel *C* which could be twisted round on a wooden peg, so that by raising or lowering the point of suspension, the position of the writing point on the drum could be adjusted without altering the load on the specimen. A short distance in front of the point of suspension, a needle *N* was driven horizontally through the middle of the lever. By threads looped round this needle the lever was connected above to a stiff spring *S* hung from another bracket on the same stand, and below to the specimen. The specimen was threaded through a pipette teat as before, and its lower end was tied to a rack and pinion by means of which it could be stretched or released. The same appa-

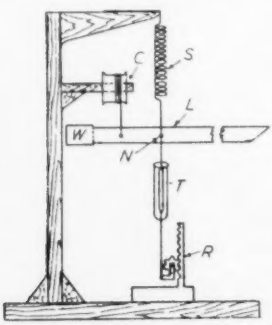


Figure 13. Adjustable isometric lever

ratus could be used for isotonic experiments if the specimen were hung in place of the spring S and a weight was freely suspended in the position T below the lever.

Description of the model system— The system used as a model was a tendon immersed in a solution of mercury-potassium iodide (Thoulet's solution). As described by SCHMIDT[34] this will give contractions very like those of a muscle. Many other swelling agents have a similar action, but it is not always those that swell the tendon most that are most effective in producing tension. Thioglycollic acid for example swells the tendon more than mercury-potassium iodide, but the tendons so swollen have a lower Young's modulus (*Figure 14*). WHEWELL and WOODS[35] have described another interesting model, which depends on the contraction of hair or wool in solutions of cuprammonium. The author has examined this too and repeated their experiments, but most of his investigations have been made on tendon, because it is more convenient. Tendons from pheasant's legs are convenient to use because they are ossified at both ends; the bony parts are not softened by the reagent and make good attachments.

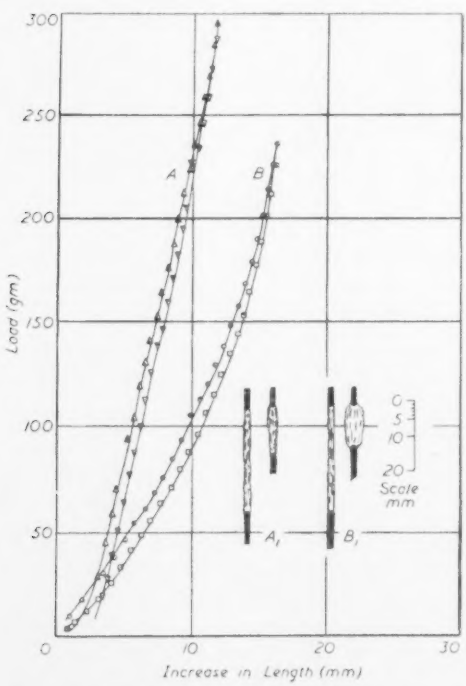


Figure 14. Load/length curves for two similar pheasant tendons, A treated with mercury potassium iodide, and B with thioglycollic acid, A_1, B_1, dimensions of the tendons before (left) and after (right) treatment, ossified parts shown black. The broadest side of the tendons is shown, but swelling is nearly isotropic in all directions at right angles to the long axis

Mercury-potassium iodide probably acts by forming relatively inert complexes with the free amino groups of the protein; in alkaline solution (Nessler's solution) it forms an insoluble precipitate with ammonia. The effect of the reagent on swollen gelatin jelly depends to some extent on the history of the jelly. One sample of cooking gelatin became opaque, but did not liquefy, if merely swollen in water before testing, whereas a jelly made by dissolving the same material in hot water and recasting into sheets disintegrated rapidly.

The way the jelly liquefies suggests that hydration of the gelatin is decreased rather than increased in the process. There is little swelling, and the individual pieces of jelly develop a thin opaque skin on their surface, which disintegrates and falls to the bottom like mist. Concentrated solutions of gelatin in the reagent have a remarkably low viscosity and show no birefringence of flow.

Tendon is dissolved very slowly in alkaline, and not at all in neutral solutions, but in both it shortens and becomes rubbery. The equilibrium load/length curve for tendon immersed in reagent of sp gr 1·3 (*c* 0·5M) is shown in *Figure 14A*; it is concave toward the load axis like that of rubber or supercontracted hair. The dynamic modulus decreases as the tendon shortens, confirming that contraction is due to a decrease rather than an increase in the strength of intermolecular forces. The critical range of frequencies for tendon contracted in a solution of sp gr 1·3 lies above 5 cycles per second (see *Figure 15*).

Fresh tendon exhibits strong birefringence, positive with respect to the fibre axis. In a fully contracted tendon the strength of the birefringence is very much decreased, but it remains weakly positive. According to KAYE and LLOYD[36] thermally contracted collagen fibres from hide become optically isotropic. ASTBURY[37] finds that the x-ray diffraction diagram of thermally contracted tendon from rat's tails shows it to be amorphous.

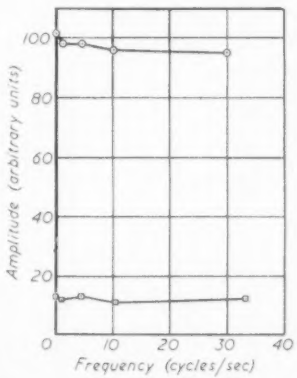


Figure 15. Variation of dynamic modulus of pheasant tendon with frequency, both fresh (lower curve), and after treatment with mercury-potassium iodide (upper curve). Tendon split into four longitudinal strands of approx. dimensions 31, 0·5 0·8 mm of which all cut away but one. Residual load 2 gm

Similar experiments were made on hair in cuprammonium and gave similar results. At a frequency of 0·1 cycle per second the dynamic modulus of human hair is decreased by about one and a half times when it is immersed in a saturated solution of copper oxide in ammonia of sp gr 0·88. This suggests that in hair too, contraction is due to a decrease in intermolecular attractive forces. Whewell and Woods themselves suggest that contraction is brought about by attractive forces exerted by adsorbed cuprammonium ions, but even if the decrease in dynamic modulus is left out of account, it seems more in agreement with the effect of cuprammonium solutions in dissolving other fibrous proteins to suppose that the adsorbed ions tend rather to decrease intermolecular attractions than to intensify them.

According to Whewell and Woods the x-ray diffraction pattern of fully shortened hair shows it to be amorphous.

Effect of concentration of reagent on load— A tendon contracting in Thoulet's solution can exert a considerable force. Starting from zero load and contracting against a stiff spring, shortening about 10 per cent, a tendon immersed in a 2M solution develops a tension of about 100 kg/sq cm on the original area. This exceeds the highest tension recorded for muscles, which is only 35 kg/sq cm (adductor of *Venus verrucosa* reported by MARCEAU[38]). To produce a tension of the same order as those developed by vertebrate striped muscles, *c* 10 kg/sq cm, a concentration of 0·5M is sufficient.

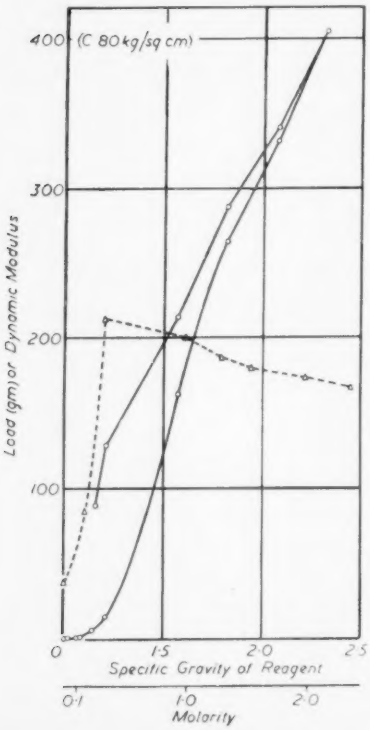


Figure 16. Variation of load at constant length (full line) and of amplitude (reciprocal of dynamic modulus) at constant load (broken curve) with concentration of reagent

The equilibrium force which can be developed at constant length (the isometric tension) increases with increasing concentration of reagent (*Figure 16*), so that the system fulfils the requirement for the working substance of a chemical engine, and by suitable variations of the concentration could be taken round a cycle analogous to Carnot's cycle, and so made to do work.

The curve for isometric tension plotted against concentration is like that for isometric tension plotted against temperatures ; a shape of curve which is typical of the load/temperature relations of imperfect rubbers. The chemical analogue of the rubbery melting point will be called the chemical melting point. The dynamic modulus at constant frequency varies with concentration as expected. There is a sudden decrease to a minimum value over a narrow range of concentrations (*Figure 16*) immediately above the chemical melting point, and then, as the concentration is raised further, the dynamic modulus increases again, because the tendon is now behaving entirely as a rubbery solid, so that the dynamic modulus increases with the static modulus.

The value of the chemical melting point is affected by tension in

the same way as that of the thermal melting point (see WÖHLISCH and DU MESNIL DE ROCHEMENT[18]) ; both are shifted towards higher concentrations or higher temperatures by applied loads.

Energy relations— When tendon is brought to contract by heating in water, heat is absorbed. As WÖHLISCH[3, 39] has pointed out, this heat can be analysed into two components, the latent heat of melting necessary to bring the tendon into a rubbery state, and extra heat absorbed during rubbery shortening. When tendon is made to contract by chemical means, the plasticizer uptake can be analysed in a similar way into latent uptake necessary to ' melt ' the tendon, and a further uptake in the range of concentrations over which tension is proportional to plasticizer content.

Adsorption of plasticizer will lead to a decrease in internal energy as attractions between plasticizer and substrate are satisfied, and will be accompanied by a corresponding emission of heat, the heat of adsorption. The effect of the adsorbed plasticizer is to reduce the attractions between molecules of the working substance, involving a further diminution of internal energy.

As long as there is an approximately linear relation between plasticizer concentration and load, adsorption of plasticizer will be proportional to the distance shortened. Hence there will be a decrease in internal energy as the tendon shortens *i.e.* $(\partial U/\partial l)_T$ is negative. If the decrease in internal energy is greater than the corresponding increase in entropy *i.e.* $-(\partial U/\partial l)_T > T(\partial S/\partial l)_T$, the process of shortening will be exothermic. If the tendon is extended by a load greater than the isometric tension, plasticizer is expelled, the internal energy increases and heat is absorbed.

The thermal coefficient of tension $(\partial K/\partial T)_l$ remains positive, because the ' melted ' tendon can act as a simple rubbery solid. This agrees with the observations of FAURÉ-FREMIET and WOELFFLIN[40] on the effect of swelling agents on the elastic properties of elastoidin.

The immediate source of the work done during contraction is an increase in the entropy of the working substance due to disorientation of the molecules, accompanied by a decrease in internal energy due to a diminution of the strength of intermolecular attractions. If we consider a complete cycle, after the tendon has returned to its original state, the only change is that the solution of plasticizer is more dilute than it was to start with. The ultimate source of energy must therefore be an increase in the entropy of the plasticizer solution resulting from dilution.

Apparent viscous-elastic effects— At concentrations above the chemical melting point the relation between tension and concentration is reversible, and a very small change in concentration will produce a

measurable change in tension ; a perceptible decrease was caused for example by adding 0·1 cc of distilled water to 3 cc of reagent. A similar change in the opposite sense follows the addition of a drop of a more concentrated solution. These changes are slow, and for a whole tendon of cross-sectional area about half a square millimetre require some ten minutes to go to completion, although a change in the slope of the tracing becomes apparent immediately.

Contracted tendon responds to sudden changes in load like a heavily damped viscous-elastic system. It regains tension after a quick release and loses it after a quick stretch (*Figure 17*). It is unlikely that these changes are due to simple elastic after effect because they are too slow. The dynamic modulus of fully contracted tendon varies little over a range of frequencies from 0·003 to 4 cycles per second (*Figure 15*), showing that adjustments due to elastic after effect ought to be complete in a matter of fractions of a second, whereas these changes proceed rapidly for several minutes and are not complete for a matter of half an hour or so. The anomaly is best explained if we suppose that tendon gives up reagent when stretched and absorbs it when released ; by le Chatelier's principle this in fact follows from the observation that tension at constant length varies directly with concentration of reagent. These changes correspond to isothermal changes in a thermal system, and as they involve diffusion of reagent in or out, the rate at which they take place should depend on the cross-sectional area of the fibre. The slow adjustments of length and tension which follow a sudden stretch sufficient to produce a temporary increase of load of about 30 gm on 500 correspond, both in rate and extent, to those produced by a change in concentration of about 30 per cent. By adding water or more concentrated reagent just before a sudden change in length, the slow changes in tension which should follow can be abolished or changed in sign (*Figure 18*).

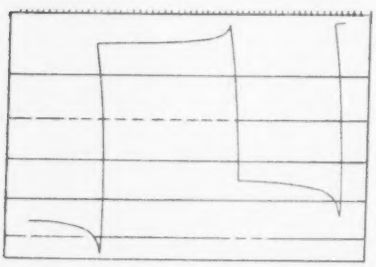


Figure 17. Effect of quick changes of length on the tension exerted by a tendon in 0·5 M Thoulet's solution; time in min. Horizontal lines represent intervals of 50 gm. Original dimensions of tendon 31 × 0·6 × 1·2 mm

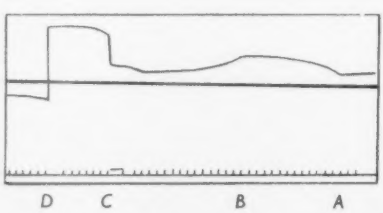


Figure 18. Tendon in equilibrium with 0·5 M Thoulet's solution. At A, *addition of 0·4 cc water to 3 cc reagent,* B *addition of 0·3 cc of 2 M solution,* C *quick release with simultaneous addition of 0·8 cc water,* D *quick stretch with addition of 0·6 cc 2 M solution of reagent, time marker in min*

If the load on a tendon is already

greater than the isometric tension, addition of reagent causes no increase. This of course follows from the principle that the equilibrium content of reagent is diminished by increasing tension. The same thing can be expressed in another way by saying that the increase in tension produced by adding reagent is greatest near the resting length of the tendon, and falls off to nothing at high loads.

If a tendon immersed in reagent, but loaded beyond its isometric tension so that it cannot contract, is given a series of quick releases, tension is regained after each, both the extent and rate of the regain

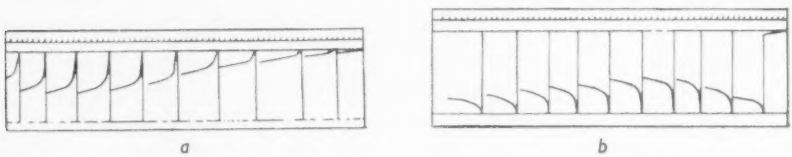


Figure 19. a Fresh tendon preloaded above its isometric tension, given a series of quick releases, each allowing 1·4 mm shortening. After each release 8 min allowed for tension to recover, zero position of lever then adjusted and another release made, until no further recovery of tension occurs; b (follows a) tendon allowed to equilibrate at zero load, and then given a series of stretches. Both records read from right to left

increasing to a maximum as the load on the tendon decreases (*Figure 19a*). A somewhat similar set of curves is obtained from a series of stretches (*Figure 19b*) starting at zero load and increasing to the full isometric tension. The shape of the curves can be explained if it is supposed that the difference between the values of the adiabatic and the isothermal or isobaric elastic moduli at any length increases with increasing plasticizer content. The increase in load caused by each successive stretch will then fall off progressively with increasing tension and with it the speed and extent of the subsequent recovery. At tensions greater than isometric, when all the plasticizer has been expelled, adiabatic and isobaric moduli must be the same, and there is no slow readjustment of tension. *Figure 20* illustrates this explanation, showing adiabatic tension/length curves superimposed on an equilibrium isobaric curve *AB*; the extent of the readjustment of load after each quick stretch or release is

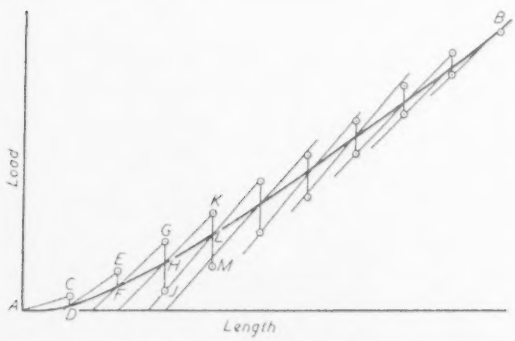


Figure 20. Variation in slope of adiabatic and 'isobaric' load/length curves at different initial lengths

taken from *Figure 19*, and the shape of the equilibrium curve from *Figure 14A*.

Rate of contraction under isotonic conditions— In isotonic experiments, with a weight freely suspended from the tendon, the rate of shortening was found to decrease with increasing load according to the equation

$$(V + 0{\cdot}086)\,(K - 150) = \text{constant}$$

where V is the rate of shortening in mm/min measured at *c* 2 per cent shortening, and K the load in gm (*Figure 21*). The rate of shortening decreases with decreasing length; *Figure 22* shows a series of curves traced from smoked drum records.

If the applied load exceeds the isometric tension, the tendon elongates instead of contracting when reagent is applied, although

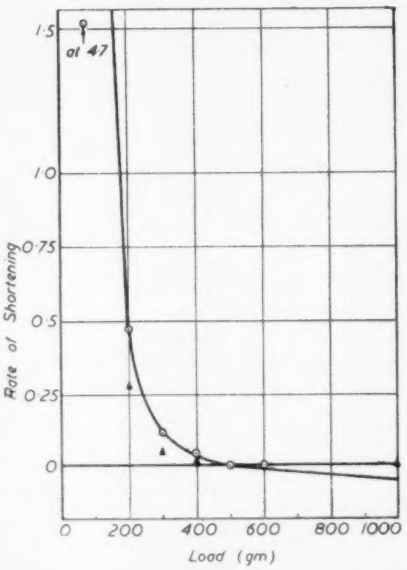


Figure 21. Variation of rate of shortening with load for tendons immersed in 1·5 M mercury-potassium iodide. Ordinates show rate of shortening in mm/min, abscissae show load in gm. Circles refer to one tendon, triangles to another

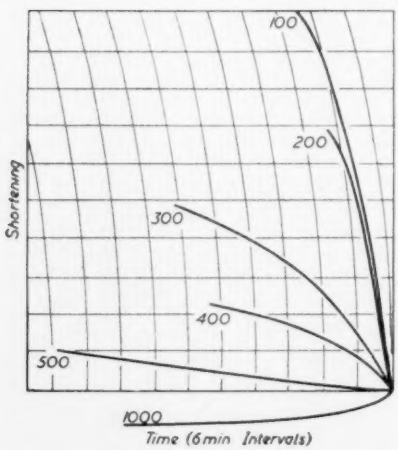


Figure 22. Superimposed tracings of contraction of a tendon in 1·5 M mercury-potassium iodide under different loads. Figures indicate load in gm, horizontal line intervals of 1 cm on drum (equivalent to 0·37 mm shortening), vertical lines, time in 6 min intervals

the rate of elongation is less than that predicted by the equation.

The relation between rate of shortening and concentration of the reagent (*Figure 24*) is like that between rate of shortening and applied load, as would be expected from the approximately linear relation between concentration and isometric tension which applies at all concentrations above the chemical melting point. There is however a very marked falling off in the rate of contraction at very high concentrations.

In general, the relations between rate of shortening, load,

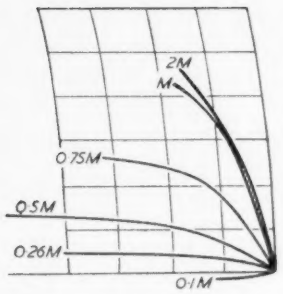


Figure 23. Superimposed tracings of the contraction of a tendon loaded with 200 gm in solutions of mercury-potassium iodide of different molar concentrations: horizontal lines indicate intervals equivalent to 1 mm shortening, vertical lines, time in 10 min intervals

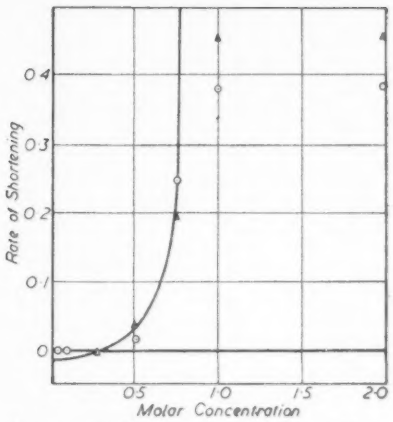


Figure 24. Variation of rate of shortening with concentration of reagent for tendon loaded with 200 gm: ordinates show rate of shortening in mm/min, abscissae, molar concentration

and concentration of reagent are best explained if we suppose that the rate of shortening depends on the rate of diffusion of reagent into the tendon; this in turn should depend on the difference between the concentration outside $[P_{ext}]$ and the equilibrium concentration inside proper to the load $[P_{int}]_K$. This can be summarized as

$$V_{pk} = b([P_{ext}] - [P_{int}]_K)$$

where b is a diffusion constant depending on the diffusion coefficient of the reagent in the working substance, and on the size of the tendon. For tendon lengthening under a load greater than the isometric tension, conditions for the diffusion of the reagent are different. Not only the direction, but the whole shape of the diffusion gradient is altered, so that the rate of lengthening cannot be expected to obey the same relations as the rate of shortening. Apart from factors affecting the diffusion of reagent, it must also be remembered that extended tendon, like resting muscle, is near the limit of its rubbery extensibility, so that it cannot extend far under any load: it is noticeable that virgin tendon, which has never been 'melted' before, extends much less than regenerated tendon, presumably because its molecules start more nearly parallel.

Relaxation and 'set'— Tendon can be set at any length by transferring it direct to water, which leaches out potassium iodide, and precipitates red mercuric iodide within the fibre. The 'set' can be reversed by washing out the mercuric iodide in potassium iodide solution. The tendon can be returned to something like its original condition by reducing the concentration of mercury slowly, or by maintaining an

excess of potassium iodide until all the mercury has been removed. A tendon treated in this way regains most of its original mechanical properties if it is kept under tension during the relaxation process, but tendon will not elongate spontaneously to any noticeable extent, so that if all the reagent is washed out and the tendon returned to distilled water it remains 'set' at a reduced length. Even under the most favourable conditions tendon never quite regains the opaque silky lustre it has when fresh, but remains translucent in patches ; the lustre returns temporarily when a regenerated tendon is stretched, but vanishes again when tension is released.

Hair in cuprammonium can be set at any length in a similar way, by transferring it to water, which precipitates blue cupric ions in the fibre. If dilute sulphuric acid (1 per cent) is added the hair gives up copper, losing its blue colour, and at the same time returns to its original length. Whewell and Woods point out that the colour is only discharged as the original crystalline arrangement is restored. In acid the hair remains much swollen, but if transferred to distilled water it will return exactly to its original mechanical properties, and regains its original dynamic modulus. This spontaneous relaxation is presumably due to recrystallization, as Whewell and Woods suggest. They observed that the x-ray diffraction fibre-diagram was lost during contraction and reappeared on relaxation ; in a stretched hair it was the β-diagram which reappeared, suggesting that the nature of the seed crystal remaining determined the type of recrystallization.

A tendon which has been set in distilled water at a length slightly below its resting length can support quite large loads, of the order of 10 kg/sq cm, without great extension. If a small concentration of plasticizer is added the tendon extends faster, presumably because the plasticizer facilitates relative movement of the chains.

At most concentrations of plasticizer, a contracted tendon will maintain tension indefinitely, at any rate for several days, without perceptible loss. In very strong solutions (c 2M mercury-potassium iodide) however, although the tension increases rapidly to a high value, it decays with time, and will show a decrease of about 1 per cent in half an hour.

MUSCLE

We may now turn to an examination of muscle. There is little evidence available to make possible a distinction between the two types of cycle. WÖHLISCH[41] has pointed out that the shape of the curve for isometric tension plotted against length (concave towards the load axis) is like the load/length curve of a rubbery material. BULL[42] has carried this comparison further, and has shown that the load/length

curve obtained from Ramsey and Street's data for isometric tensions developed by a single fibre of frog muscle at different initial lengths is closely similar to that for rubber.

Optical changes during activity— Both types of cycle require that the working substance should consist of long chain molecules which shorten by curling or folding, and it has been shown that muscle proteins fulfil this condition. In resting muscle, both form and intrinsic birefringence are strong, showing that a considerable proportion of the proteins are crystalline; this is confirmed by the x-ray diffraction diagram. The form birefringence is positive with respect to the long axis of the muscle fibres, showing that the crystalline micelles are elongate bodies lying with their long axes parallel to the length of the fibres; this again is confirmed by the form of the x-ray diffraction diagram. When a muscle shortens, the birefringence decreases. It is probable that both form and intrinsic birefringence are affected, suggesting that the proportion of crystalline material is diminished (see review by FISCHER[43]). In the x-ray diffraction diagram the spots become slightly more diffuse and widen into arcs, but there is no change in their orientation, showing that although some of the crystalline material assumes a less orderly arrangement, the remaining crystallites keep their original alignment (ASTBURY[6]). The spacing of the interference spots in the diagram of resting muscle is like that given by unstretched hair (α-keratin), suggesting that the protein chains in the micelles are folded. When the muscle contracts they presumably fold closer still, but not in a way sufficiently regular to yield a reflection in the x-ray diagram. A resting muscle frozen in liquid air breaks into fibrous splinters when struck with a hammer. A contracted muscle gives a glassy fracture (MEYER[23]).

The smallest structural units visible under the electron microscope in material from macerated muscles are fibrils about 150 Å in diameter. In material from muscles fixed in the contracted state they appear to be shorter and thicker, but show no signs of bending or folding (HALL, JAKUS and SCHMITT[44]).

Temperature effects and heat exchanges— All these observations are consistent with either type of cycle. Contraction is evidently due to the folding or curling up of chain molecules, but there is nothing to show why they fold. Many people have tried to decide the question by measuring the temperature coefficient of isometric tension $(\partial K/\partial T)_l$ in active or resting muscle, supposing a positive coefficient to be diagnostic of an entropy cycle. This assumption would be justified if muscle were a heat engine, but for a chemical engine, although it probably is true, the connection is not at all an obvious one. For active muscle the position is further complicated by the effect of

temperature changes on the rate of the accompanying chemical reactions ; but bearing all these limitations in mind the evidence is perhaps more easily explained on the hypothesis of an entropy cycle than any other.

HARTREE and HILL[45] found that both resting and active muscle gave out heat when stretched, and according to RAMSEY, quoted by BULL[42], the thermal coefficient of isometric tension in single fibres of frog muscle is positive. For resting muscle, MEYER and PICKEN[46] have measured $(\partial K/\partial T)_l$ by heating in a water bath, and find it positive ; their results have been confirmed, as to the sign of the coefficient at any rate, by RENK and WÖHLISCH[47], and by BUCHTAL, KAISER and KNAPPEIS.[48] Observations of the heat exchanges accompanying activity are difficult to interpret because they represent the sum of several physical and chemical processes, each absorbing or emitting heat in unknown quantities : any hypothesis of the mechanism of muscular contraction must be in accordance with the observed facts, but the facts as they stand do not immediately exclude either type of cycle.

A crucial experiment— The most promising method remaining to decide between the two kinds of cycle seemed to be to compare the elastic properties and viscosity of muscle in both the resting and the active states. For this, the alternating stress machine already described was used.

The muscle used was the anterior byssal retractor of *Mytilus edulis*, which is a flat strip of tissue about 30 mm long by 3 mm wide. The structure and histology have been described by FLETCHER[49] and by NIEUWENHOVEN[50]. The fibres run approximately parallel and extend from one end of the muscle to the other. The muscles were tied with a linen thread at both ends before they were cut out ; one thread was knotted close to the muscle so that it could easily be caught in the hook on the lower edge of the spring on the apparatus, and the other was left long, with a loop at the end to which was attached a 2 gm weight. The lower thread could be gripped between the muscle and the weight by a clamp C (*Figure 12*) which held it fast in one position. The greatest load imposed on the muscle by the spring when stretched by the eccentric was 10 gm. After dissection muscles were allowed to rest for at least 3 hours suspended in aerated sea water at 0° with a 2 gm weight attached ; best results were obtained by leaving them overnight.

The muscle could be stimulated through two electrodes, connected to the 220 v AC mains through suitable resistances. The upper electrode was a glass tube into which was sealed a platinum wire connected to the AC circuit. The position of the tube could be

adjusted, and it was arranged to deliver a fine stream of sea water on to the upper end of the muscle. The lower electrode was a strip of flannel soaked in sea water which was wrapped round the lower end of the muscle, and made contact with a silver plate connected to the AC circuit.

The apparatus differed from that used by WINTON[33] mainly in that it incorporated a quick-acting clamp, so that the muscle could be allowed to shorten isotonically before it was tested.

The amplitude of the response given by a resting muscle depended on its initial length, being smaller the longer the muscle. This is a factor which it is difficult to control, because muscles are subjected to a variable amount of stretching during dissection. The most consistent results were obtained by stretching the muscle before each experiment, applying a load of about 20 gm by hand and maintaining it for one minute, by which time the first rapid elongation had ceased. This treatment had an effect, which persisted indefinitely, in decreasing the amplitude given by a resting muscle when subsequently tested under a residual load of 2 gm. The effect was abolished by contraction of the muscle, and it was found that the properties of active muscle were not affected by previous stretching.

Apart from the persistent effect of stretching, the dynamic modulus of both resting and active muscle at any frequency decreases with increasing tension. This effect does not persist when the tension is released, and is simply a consequence of the fact that the load/length curve for both active and resting muscle is concave towards the load axis. Similar effects have been described in detail by BUCHTAL, KAISER and KNAPPEIS[48] in single fibres from frog muscle. The magnitude of the effect can be seen from *Figure 25*.

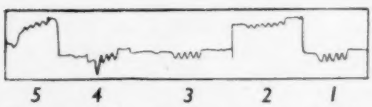


Figure 25. Effect of tension on the dynamic modulus of Mytilus *anterior byssal retractor muscle, frequency 0·1 cycle/sec. From right to left: 1 resting unstretched, 2 resting stretched, 3 resting after stretching, 4 stimulated by AC, isotonic, 5 same, isometric*

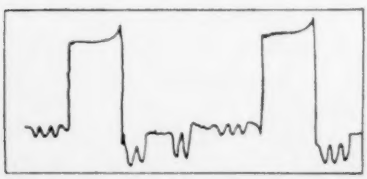


Figure 26. Response of Mytilus *anterior byssal retractor muscle to oscillating tension at 0·1 cycle/sec. From right to left: 1 resting unstretched, 2 stretch, 3 resting after stretch, 4 stimulated by AC, isotonic, 5 resting after stimulation, 6 stretch, 7 resting after stretch*

In *Figure 26* are shown tracings obtained in response to an oscillating tension at 0·1 cycles per second from a *Mytilus* anterior byssal retractor. It will be seen that there is an increase in the amplitude of the response (decrease of dynamic modulus) when the muscle is active, and that immediately after stimulation the amplitude is large and the

slope of the base steep, so steep in fact that it is difficult to obtain reliable readings. After stretching, the muscle returns to its original condition. *Figure 27*, taken at a lower frequency, shows a larger increase in amplitude on stimulation, and also exhibits, better than *Figure 26*, the characteristically steep slope of the base line in active muscle.

Figure 27. As Figure 26, but frequency 0·05 cycles/sec : 1 resting after stretch, 2 stimulated by AC isotonic

Variation of amplitude with frequency was found to be very much as described by Winton. The amplitude given by active muscle drops rapidly at a frequency between 1 and 0·2 cycles per second, and above this range shows only a slow decrease with increasing frequency. The amplitude given by active muscle reaches a maximum at about 0·02 cycles per second, while that of resting muscle is still increasing slowly at this frequency. At frequencies much above 1 cycle per second the amplitude given by active muscle is less than that given by resting muscle (*Figure 28*).

A similar series of experiments were tried on frog's muscle, pairs of sartorius muscles, and single gastrocnemius preparations. The sartorius muscles were suspended above by a thread passing through the bone of the acetabulum and below by a thread tied round short pieces of femur which were left attached. The gastrocnemius preparations gave similar results to the sartorius muscles, and were used for all later experiments because they were more convenient. Both muscles were stimulated direct by AC stimuli of intensity just sufficient to cause a full response. The electrodes were the same as for experiments on *Mytilus* muscles, except that frog Ringer was used instead of sea water.

The results are like those obtained with *Mytilus*. *Figure 29* shows a tracing from a gastrocnemius at 1·0 cycle per second ; as with *Mytilus* muscle there is a pronounced decrease in the dynamic modulus on stimulation. Variation of dynamic modulus with frequency was investigated, but it was found that the critical range of frequencies for active muscle lay above the highest frequency of which the apparatus was capable (4 cycles per second). At frequencies from 4 to 0·2 cycles per second there is little change of amplitude with frequency for either active or resting muscle (*Figure 30*). Passive stretching has the same effect on the subsequent behaviour of resting muscle as it does in *Mytilus*.

This picture can be completed by reference to the work of BUCHTAL, KAISER and KNAPPEIS[48] who have measured the dynamic modulus of single fibres from the semitendinosus muscle of the frog over a wide range of frequencies. They find that the lower limit of the critical range for active fibres lies between 2 and 6 cycles per second. The dynamic modulus of resting fibres was still equal to twice the static modulus at the lowest frequency investigated, namely 0·5 cycles per second. As in *Mytilus* muscle, the dynamic modulus of active muscle is slightly greater than that of resting muscle at all frequencies above the critical range of active muscle. Assuming that the behaviour of semitendinosus fibres is not very different from that of gastrocnemius muscles, this gives a picture of the variation of dynamic modulus with frequency for frog muscle like that for *Mytilus* muscle, except that the critical range is at a higher frequency.

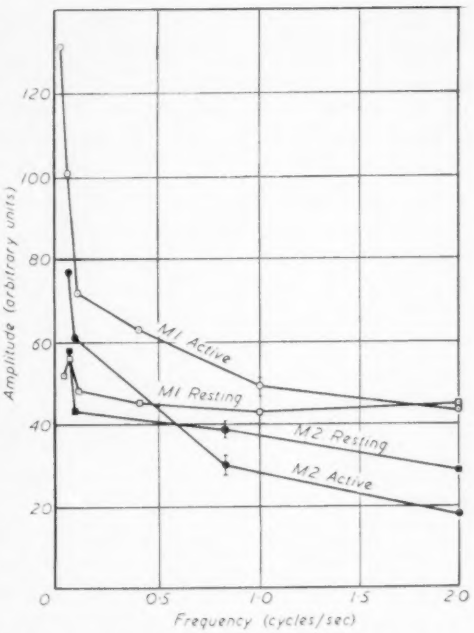


Figure 28. Variation of dynamic modulus of Mytilus *byssal retractor muscles with frequency: 'active' muscles stimulated by minimal AC shocks: M1 M2 muscles from two different animals*

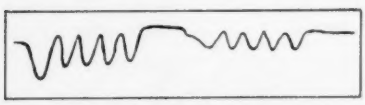


Figure 29. Response of a frog gastrocnemius to oscillating tension at a frequency of 1 cycles/sec. From right to left: resting after stretch and during stimulation by AC

'*Potential*' *hypothesis*— In its simplest form, the 'potential' hypothesis leads one to expect an increase in dynamic modulus at all frequencies when a muscle contracts; it can only be adapted to explain a decrease by making it more complicated. A decrease in dynamic modulus on stimulation could be explained for example if all the load were taken by a relatively inextensible sarcolemma or by connective tissue in the resting muscle, and only came on to the contractile mechanism when the muscle shortened.

There is no doubt that in a muscle such as the frog gastrocnemius there is a good deal of connective tissue in parallel with the contractile mechanism, and the importance of the sarcolemma is suggested by the observations of RAMSEY and STREET[51] on single fibres from frog's muscle. They found that the elastic modulus of a whole fibre was not affected if the cytoplasm was damaged locally, even to the point of being completely interrupted, provided the sarcolemma was intact. This result seems to prove too much however ; even a stocking full of sand would become more extensible if the sand was arranged to leave a gap in the middle. SICHEL[52] has measured the extension of short lengths near the damaged part, marked by specks of carbon, and he finds that where the cytoplasm is interrupted the empty sarcolemma is on the average twice as extensible as an intact fibre. The cytoplasm near the injury coagulates, and seems to become considerably less extensible, as a length containing coagulated lumps as well as empty lengths of sarcolemma may have nearly the same overall extensibility as an intact fibre. RAMSEY[53] has criticized Sichel's work on the ground that extension may have been due to tearing of the fibres by the needles used for attachment, but this objection is not valid as the gauge length was not in fact in contact with the needles at all.

In *Mytilus* muscles the sarcolemma is thin, and there is no connective tissue visible in cross section. If the muscle is loaded by a weight of more than about 5 gm it extends indefinitely until it breaks, so that if the sarcolemma does take the load it must be plastic, like the material it encloses. The viscosity of resting muscle is certainly not so high as to be beyond the capabilities of the muscle proteins; the viscosity of muscle in contracture, or in rigor in the shortened configuration, may be higher than that of resting muscle.

For *Mytilus* muscle at any rate, it must then be admitted that the mechanical properties of muscle are essentially those of the muscle proteins, and if so the viscosity of the proteins must decrease on contraction. To reconcile this conclusion with any form of ' potential ' hypothesis requires further elaboration in the molecular mechanism proposed. ASTBURY[54] supposes that the protein chains, or rather grids formed of aggregates of chains, are subjected to two sets of forces ; strong lateral attractions which hold them together in the folded configuration, and another, weaker, set of forces which tend to fold them more tightly still. On stimulation the former are abolished and the chains fold more tightly. This hypothesis is open to the general objections discussed on p 232, and it fails to explain why the load/length curve of active, as of resting, muscle should be concave toward the load axis, or why the dynamic modulus of active muscle

should be higher than that of resting muscle at frequencies above the critical range but less at frequencies below it.

BUCHTAL[55] explains this last difficulty by supposing that the muscle is a diphasic system ; a continuous but porous contractile element, impregnated with a viscous fluid. This is also the hypothesis on which WINTON[33] interpreted his results. It will account for most of the facts, but like all spring and dashpot hypotheses, it fails to explain why an increase of intermolecular forces in the contractile element should be accompanied by a decrease in the viscosity of the fluid part.

'Kinetic' hypothesis— The 'kinetic' hypothesis seems to offer a simpler and more complete explanation than any other. The decrease in viscosity on stimulation, and the variation of dynamic modulus with frequency are as expected, both for *Mytilus* and frog muscle. The reversible increase in dynamic modulus with increasing tension in both active and resting muscle is typical of rubbery materials, as has been pointed out by WÖHLISCH[41] and by BULL[42]. The only detail not immediately accounted for is the increase in dynamic modulus at frequencies above the critical range observed on stimulating both *Mytilus* and frog's muscle. The change is small compared with the decrease observable within the critical range, and although on the kinetic hypothesis it is not immediately concerned in the production of tension, it is an interesting clue to the changes in the muscle proteins. At frequencies above the critical range the rotation and uncurling of chains is largely suppressed, and extension is due to stretching of the chains themselves, so that changes in the dynamic modulus must be a consequence of changes in the stiffness of the chains rather than of differences in the forces between them. If plasticizer molecules become attached to several consecutive polar groups along the protein molecules, it seems reasonable to suppose that they may reinforce the chains and make them less extensible.

According to the kinetic hypothesis, the sequence of events in a stimulated muscle is as follows. At rest the proteins are in a partially crystalline condition, aligned parallel to the long axis of the muscle. Even in the non-crystalline parts, intermolecular attractions are strong, so that no tension is exerted, and the dynamic modulus is high. Stimulation causes the liberation of a plasticizer, as a working hypothesis let us suppose adenosine triphosphate ATP, which is strongly adsorbed on to the proteins and reduces the attractive forces between chains ; as the strength of intermolecular attractions decreases, the tension increases. The dynamic modulus at frequencies near the critical range of resting muscle decreases as the critical range shifts toward higher frequencies.

In the resting muscle we must suppose that adenosine triphosphate

is present, but for some reason is not available to the muscle proteins ; ions necessary for its action may be lacking, or the ATP itself may be bound at a polar membrane. Stimulation by alternating current breaks down the polar membrane, or otherwise sets free ATP, which is adsorbed on to the proteins. Almost immediately after being adsorbed ATP is decomposed by an enzyme closely associated with the proteins (MEYERHOF[56]) into products which are no longer capable of remaining adsorbed or of acting as plasticizers. When a muscle is in a state of sustained activity, ATP is continuously resynthesized by a process involving, ultimately, the oxidation of glucose ; as each molecule is broken down and comes off the protein it is replaced by another, so that the concentration of adsorbed ATP is maintained. When the stimulus ceases, the polar membrane is reconstituted and has the effect of immobilizing a proportion of the resynthesized ATP, so that the concentration falls.

This cycle differs from the model described above in that, during extension, plasticizer is destroyed by a chemical reaction instead of being removed by dilution ; before the cycle can be repeated the plasticizer must be resynthesized. As with tendons in mercury-potassium iodide, the immediate source of energy for contraction is an increase in the entropy of the protein chains. If we consider a complete cycle, the ultimate source of energy is a decrease in the chemical (potential) energy of the ATP. If destruction of the plasticizer is complete, the difference between the two working concentrations may be large, so that the efficiency of the cycle may be correspondingly high.

If the activation energy for adsorption is higher in the crystalline than in the amorphous regions of the protein, ATP will tend to accumulate, as the concentration falls, in parts of the chains which have not yet crystallized, and by maintaining free mobility of the non-crystalline parts it will promote rapid crystallization. The rate of crystallization will be increased by any applied force that tends to align the chains parallel.

If contraction is due to a weakening of attractions between chains, relative movement of chains will be facilitated during activity, and molecular orientation will deteriorate. As the angular distribution approaches a condition in which there are an equal number of links in every direction, the tension will decrease to nothing. Orientation can only be restored by a renewal of the process of crystallization.

Most muscles possess some power of regaining their resting length without any applied load, although its extent seems to have been much exaggerated. If this power is due to crystallization, it must involve seed crystals in the form of crystalline micelles or active surfaces which have survived the process of contraction. If not

opposed by any load, contraction may go so far that all the seed crystals are melted, and relaxation becomes impossible unless the muscle is stretched. This would seem to be the explanation of the delta state observed in single fibres by RAMSEY and STREET[51] or of the 'waxy degeneration' of earlier authors (see review by HØNCKE[57]).

VISCOUS-ELASTIC PROPERTIES AND HEAT EXCHANGES OF MUSCLE

Muscles appear to possess a fairly high internal viscosity, and they exhibit what seems to be typical viscous-elastic behaviour in response to sudden changes of load. Hill, Fenn, and others studied the effect that internal viscosity might be expected to have on the rate at which muscles could do mechanical work, and they found that in a general way the mechanical effects were consistent with the idea that the contractile elements worked against a viscous resistance. The faster a muscle shortened, the less mechanical work it could do, presumably because a higher proportion of the available energy was dissipated in overcoming viscosity. The relation between speed and load was found not to be linear, but this is not in itself a fatal objection to the theory, being a characteristic shared with many other high polymers.

Measurements of the heat exchanges of contracting muscle did not agree with this simple theory however. In 1923, FENN[58] showed that the total energy production (heat + work) of a muscle was greater when it was allowed to shorten than when it was made to contract isometrically, and in 1935 FENN and MARSH[59] showed that it was in fact impossible to account for both the heat exchanges and the mechanical properties of muscle on the basis of a simple viscous-elastic model. They suggested that the observed relation between rate of contraction and the amount of work done depended on a fundamental property of the contractile mechanism, which developed 'shortening energy' less rapidly at low loads. HILL[60-64] has developed this theory further, and has made it more exact, basing his conclusions on the results of very accurate measurements of temperature changes in active muscle.

HILL[61] analyses the heat production of an isotonic muscle twitch into two components; an 'activation heat' which is approximately constant under all conditions, and 'heat of shortening'. The rate at which heat of shortening is evolved varies inversely with the load, but the total amount depends only on the distance through which the muscle shortens. He can find no evidence[62] for the existence of any 'heat of relaxation'. In a sustained tetanic contraction there is a steady evolution of 'maintenance heat'.

To interpret these observations in the light of our findings on the behaviour of a tendon in Thoulet's solution, we must suppose that the heat given out represents heat of adsorption of plasticizer. 'Activation heat' then corresponds to the 'latent uptake' and 'heat of shortening' to the extra uptake associated with shortening. The existence of maintenance heat is due on this hypothesis to a continuous cycle of decomposition, removal, resynthesis, and re-absorption of plasticizer. According to MEYERHOF[56] the only important source of heat is the splitting of ATP because the heat exchanges of the other reactions in the glycolysis balance out. The total turnover of plasticizer in this cycle will depend on the amount adsorbed at any one time, and should be increased when more is taken up, so that the rate of production of 'maintenance heat' should be greatest at low loads. When the stimulus ceases, plasticizer is progressively removed from circulation as it is resynthesized, so that the cycle comes to an end and the production of maintenance heat shows an immediate decrease.

In a tendon contracted in Thoulet's solution, an increase in tension expels plasticizer, and should be accompanied by a heat exchange corresponding to the heat of adsorption, but of the reverse sign. This effect is not to be expected in such a simple form in muscle, because plasticizer is not removed by dilution but by chemical destruction. The rate at which plasticizer is destroyed is not necessarily affected by an increase in tension, but if tension is increased the equilibrium content of plasticizer will be diminished, and plasticizer will be taken up as it is resynthesized in reduced amount, according to the new equilibrium conditions. The final effect of an increase in tension will thus be the same as in the tendon model ; the amount of adsorbed plasticizer will be decreased, and the strength of the intermolecular attractions increased, thus increasing the internal energy, and resulting in an absorption of heat. HILL's measurements[60] showed a decrease in the 'maintenance heat' when a muscle lengthened slowly under a load just too great for it to support, and he has since found[65] that some of the work done on the muscle does not reappear as heat.

The mechanical effects of sudden change of length in an active muscle, as described by GASSER and HILL[66], are like those which follow similar changes in a contracted tendon. A slow stretch causes a slow rise in tension direct to the new equilibrium value. A rather faster stretch, complete in a matter of a tenth of a second, will cause an increase in tension to a value above the new equilibrium, followed by a slow decline to the final value. This may be accounted for as an adiabatic stretch followed by a decrease in tension as the plasticizer content adjusts to a lower value. For still faster stretches a new

factor comes in ; the muscle appears to yield ' like a wire stretched beyond its elastic limit '. The tension at first rises, then sinks back to its original value or even below (SEEMANN[67]) and finally recovers to the new equilibrium, tension increasing with time in the same way as it does in a twitch. This behaviour can be explained if it is supposed that during rapid extension, the tension rises above the equilibrium tension for the new length. It seems probable that it does, because the viscous resistance of an ' imperfect rubber ' to rapid extension may involve large forces derived from potential energy changes (see pp 225, 226). If the tension rises high enough, the plasticizer content will tend to fall during extension, and may reach a level below that appropriate to the new length. Tension at the new length would then start below the equilibrium value, and could only increase as more plasticizer diffused in.

This effect can only be shown by systems capable of rapid adjustment of plasticizer content, so that a sufficient loss of plasticizer can take place during the very short period for which the tension is high. In a muscle, the controlling factor will be the rate at which plasticizer is destroyed, which is in general high. For a system relying entirely on diffusion, the cross section of the fibres would have to be very small in order that diffusion might be fast enough. It is not surprising therefore that the transitory drop in tension could not be produced in tendon by any rate of stretching the author could devise.

The muscles of different animals differ in their response to a quick stretch. As a general rule, very slow muscles such as those of *Holothuria* (HILL[68]), or the lantern muscles of *Echinus* (LEVIN and WYMAN[69]) do not show the transitory drop in tension, while rather faster muscles such as frog's sartorius (GASSER and HILL[66]), tortoise biceps cruris or the claw muscles of *Maia* (LEVIN and WYMAN[69]), do show it. Dogfish jaw muscles, examined by Levin and Wyman, seem to be an exception in that they are reasonably fast, and yet do not show the effect. These differences may perhaps be explained as resulting from differences in the rate at which plasticizer is broken down. From a functional point of view one would expect it to be broken down more slowly in slow muscles than in fast muscles capable of rapid relaxation. HUMPHREY[70] has examined myosin from *Maia* dogfish, and a clam, *Mya*, and finds that their activity in decomposing ATP decreases in that order.

HILL[60] observes that the yielding of muscles under high loads is accompanied by the evolution of heat nearly equivalent to the work done : a similar evolution of heat accompanies the forcible extension of a muscle after the stimulus has ceased. In this the muscle is behaving like a simple rubbery solid, as would be expected if all the

plasticizer had been expelled, so that the situation was no longer complicated by negative ' heat of desorption '.

Rate of shortening ; isobaric changes— The heat of shortening, as we have seen, is proportional to the distance x through which the muscle shortens. If the load on the muscle is K, the total energy in excess of that liberated in an isometric contraction is $K(a + x)$, where a is a constant defining the ' shortening heat ' per unit distance. The rate at which extra energy is liberated $(K + a)dx/dt$ is found to be proportional to the amount by which the load falls short of the isometric load, K_0. Taking $dx/dt = V$

$$(K + a)V = b(K_0 - K)$$

where b is a constant defining the increase in the rate of energy liberation per unit decrease in load. This equation can also be written as

$$(K + a)(V + b) = (K_0 + a)b = \text{constant}$$

From these equations HILL[60] shows that both the relation between rate of shortening and work done, and the rate of increase of tension in an isometric contraction can be deduced. RAMSEY[53] has pointed out that by themselves the equations do not set any limit to the extent of shortening possible, nor imply that the rate of shortening decreases with decreasing length, as it clearly does. These omissions do not however affect the application of the equations under the conditions originally specified ; small changes of length, involving no great changes of load.

We have already seen that tendon contracting in a solution of mercury-potassium iodide shows a similar relation between load and rate of shortening. In fact this type of relation follows if the rate of contraction depends on the rate at which plasticizer can diffuse into the working substance, as already discussed.

The value of Hill's constant a is an expression of the specific heat of absorption of the plasticizer on to the proteins. Hill's constant b seems to be a hybrid, whose value depends both on the rate of diffusion of plasticizer and on the rate at which it is destroyed ; the high temperature coefficient $Q_{10} = 2{\cdot}05$ is more suggestive of a chemical reaction rate than of a diffusion constant. Conditions in the tendon model are simpler in this respect.

When extending under loads greater than the isometric tension, muscle no longer obeys the same relations between speed and load as it does when shortening (KATZ[71]). In this it agrees with tendon in mercury-potassium iodide, and much the same explanations apply, except that in muscle we are not concerned with the loss of plasticizer by diffusion, but with the rate at which it is destroyed by chemical action.

If the rate of contraction is determined by the rate of diffusion of a plasticizer, we should expect it to vary approximately with the square of the fibre radius (HILL[72]). On this relative basis, *Mytilus* byssal retractor is slower than vertebrate striped muscle, but faster than tendon in 0·5 M mercury-potassium iodide. Approximate figures are given in *Table I.*

Table I. Rate of Contraction of Different Muscles

	Radius r mm	*Time t to reach half isometric tension (sec)*	t/r^2
Tendon in 0·5M mercury potassium iodide ..	0·5	1,800	7,200
Mytilus *byssal retractor* [49]	0·005	0·5	2,000
Frog sartorius[72]	0·05	0·2	80
Kitten internal rectus [72] ..	0·01	0·004	40

Adiabatic viscous-elastic effects— Apart from changes involving gain or loss of plasticizer comparable to isothermal changes in a thermal system, muscle also exhibits true adiabatic viscous-elastic behaviour of the type shown by monophasic systems (p 224). The adiabatic changes are faster than isobaric changes, and it is with them that we are mainly concerned in the alternating stress measurements described above. The distinction between the two kinds of strain is well shown by frog's muscle. Careful measurements have been made by PETIT[73] of the rate at which an active frog sartorius responds to a sudden change of load, and he finds that the adjustment requires a time of the order of a second. The author's curves for the variation of dynamic modulus with frequency show the expected small change in this region (*Figure 30*). These are presumably isobaric changes. The critical frequency of frog's muscle, at which there are much greater changes in dynamic modulus, lies at frequencies of about 8 per second ; this is presumably associated with adiabatic changes.

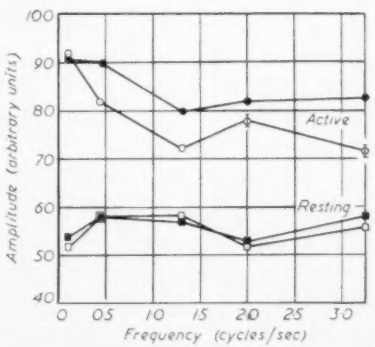


Figure 30. Variation of dynamic modulus of frog gastrocnemius muscle with frequency : two pairs of curves for a pair of muscles from the same frog

In this connection it is relevant to examine the findings of GASSER and HILL[66]. They measured the effect of a muscle in damping out the vibrations of a steel spring, and found that the damping coefficient of the active muscle was sixteen times greater than that of a resting muscle. In discussing their results they assume that the increase in damping coefficient must correspond to an increase in viscosity, but this is only true over a limited range of viscosities. If the viscosity is

so high that the relaxation time is long compared with the duration of one cycle of oscillation, an increase leads to a decrease in damping, because the material behaves more like a glass than a viscous liquid. Viscosities high enough to show this effect are common, indeed usual, among high polymers; it has been found for example that rubber antivibration mountings in aircraft become less effective at very low temperatures because of the increase in the viscosity of the rubber. To judge from Buchtal's results on single fibres from the semitendinosus muscle, the frequencies at which Gasser and Hill worked (6 to 8 cycles per second) are at the upper limit of the range over which a decrease in dynamic modulus might be expected on stimulation. It is significant that HOGBEN and PINHEY[74], who repeated these experiments at a slightly higher frequency range (9·7 to 14·7 cycles per second) could not find any difference between active and resting muscle.

Fast and slow muscles— On the potential hypothesis there is no reason why a muscle should not maintain a tension indefinitely, but from the kinetic hypothesis it appears that the same change that produces tension must also promote its decay, because the low internal viscosity associated with a high tension promotes the decay of the molecular orientation on which the tension depends.

This relation is well shown in inanimate systems in both chemical and thermal cycles. Tendons in mercury-potassium iodide will maintain a tension for weeks at low concentrations of the reagent, but at very high concentrations (2M and upward) the tension, though great at first, decays with time. Analogous behaviour is shown by nylon filaments in hot water; tension increases with increasing temperature above the rubbery melting point, but at temperatures above 80°C the decay of tension at constant length through plastic flow is so fast that the apparent value of the thermal coefficient of tension falls off (see *Figure 8*).

If a muscle retains a comparatively high internal viscosity (high value of *m*, p 224) during activity, the load developed will be small, but will decay slowly, and the muscle will oppose sudden applied loads by a large viscous resistance. This condition might be brought about either by having a low concentration of plasticizer, or by the presence of intermolecular linkages of a type not affected by the plasticizer. Muscles whose function it is to give tonic contractions probably have stronger intermolecular linkages than fast muscles. This is shown by differences in their dynamic moduli. For example, the lower limit of the critical range of frequencies for active *Mytilus* byssal retractor muscles is about 0·5 cycles per second, whereas the corresponding figure for single fibres from the frog's semitendinosus is between 2 and 6 cycles per second (BUCHTAL, KAISER and

KNAPPEIS[48]), and for whole frog gastrocnemius more than 2 cycles per second.

If full advantage is to be taken of the possibilities of a slow muscle for the economical support of loads, the rate at which plasticizer is broken down must be slow. It must correspond in fact to the rate at which tension would decay by viscous flow if plasticizer were not destroyed at all. Differences in the way in which fast and slow muscles yield under high loads suggest that this is so.

The study of their mechanical properties suggests that smooth and striated muscle fibres differ in degree rather than in kind, and that the striations are probably not a necessary part of the contractile mechanism. If contraction is caused by a substance diffusing into the middle of the fibre from the outside, the outer layers will contract first, and may expend some of their energy in shortening the core by making it buckle. In elastoidin fibres plunged into mercury-potassium iodide, shear forces set up in this way between the outside and the inside may be strong enough to splinter the fibre (SCHMIDT[34]). It seems possible that the function of cross striation may be to eliminate tension gradients across the fibres ; instead of diffusing in from the outside, the plasticizer diffuses up and down, becoming concentrated in the isotropic discs during rest and diffusing into the anisotropic discs on stimulation. If the plasticizer concerned is indeed adenosine triphosphate, this idea receives some support from the work of CASPERSSON and THORELL[75] and of ENGSTRÖM[76] who have shown that both nucleotides and phosphate are concentrated in the isotropic bands in resting muscle.

Tonus— Even slow muscles are not capable of sustaining loads indefinitely, in the way that may be necessary to maintain postural tone, and various special mechanisms have been evolved to fulfil this function.

In vertebrates postural tone is maintained *in vivo* by rapid asynchronous alternations of rest and activity in individual fibres, allowing to each intermittent periods of rest, while maintaining the mean tension on the muscle as a whole. *In vitro*, and probably *in vivo* too, the maintenance of a tetanic contraction is facilitated by fatigue, which tends to increase intermolecular attractions, thereby diminishing the maximum tension but extending the length of time for which it can be sustained. The effect is probably due to a rapid reduction in the concentration of available ATP resulting from a decrease in the rate of resynthesis.

HARTREE and HILL[45] showed that the rate of heat production associated with the maintenance of a steady tension in frog sartorius muscle was decreased after previous activity. BRONK[77] and FENG[78]

repeated and extended these results, and showed that in a prolonged tetanus the heat production to maintain a given tension might be decreased by a factor of five. Bronk[79] described a similar effect from the adductors of the claw of a crustacean *Maia* and Bozler[80] has found an even greater effect (decrease in the heat production required to maintain a given tension of between six to ten times) for the buccal retractor of *Helix*. The effects of fatigue seem to be most conspicuous in thick muscles, in which deficiency of oxygen and accumulation of metabolites would be expected to be most severe. In single fibres the effects of fatigue are so slight as to be difficult to detect (Ramsey[53]).

The author has investigated these effects on the alternating stress machine and found, as might be expected, that the slowing of the muscle associated with the increased economy with which it can maintain a tension is associated with an increase in dynamic modulus and a decrease in plastic flow. *Figure 31* shows tracings obtained in a series of short tetani from a frog gastrocnemius. It may be compared with the records given by Hill[81], which show a marked decrease in the rate of relaxation for successive twitches of a frog gastrocnemius. *Mytilus* muscle shows a similar effect. A muscle from an animal in bad condition, or one which has been fatigued by repeated stimulation, contracts more slowly than a healthy muscle and relaxes much more slowly. The maximum tension that it can exert in an isometric twitch is reduced, but the period for which tension can be sustained is increased. Any muscle can be reduced to this condition by repeated stimulation over a period of two or three minutes, particularly if not well irrigated ; the muscles of animals which have been kept for some months in the aquarium are generally like this. At frequencies below the critical range their dynamic modulus is high during activity, in extreme cases even higher than their resting modulus, and after the cessation of the stimulation may take up to 30 minutes instead of about 1 minute to return to a typical resting value. The actual

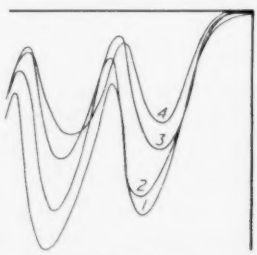


Figure 31. Superimposed tracings obtained from a frog gastrocnemius during four successive short tetani, in order numbered, frequency 0·1 cycles/sec. Note decreasing amplitude and decreasing slope of the base line

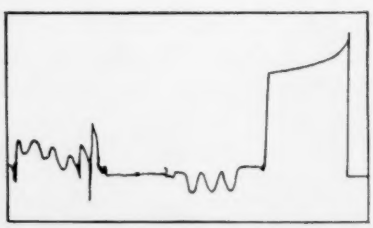


Figure 32. Tracing from a fatigued Mytilus *byssal retractor muscle, frequency 0·05 cycles/sec. From right to left : resting after stretch, and stimulated by AC, isotonic. The two peaks at the right of the second record show where the tension has been allowed to develop, and has been released before testing. Note small amplitude in active muscle, imposed on a rising base line*

process of contraction may extend over a period of about 30 seconds, so that it is possible to impose several cycles of alternating tension before it is complete. If this is done a tracing is obtained in which the usual variations in length are superimposed on a base line showing increasing tension (*Figure 32*). As this kind of fatigue is best developed in muscles which are not irrigated during stimulation, the paradoxical situation arises that a muscle can support a load for a longer time under bad conditions than under good.

Except that it is reversible, this condition is like *rigor mortis* as described by BATE SMITH and BENDALL.[82] These authors find that the development of *rigor* depends on the absence of ATP rather than on the presence of lactic acid. The development of *rigor* corresponds in time with the disappearance of ATP, and is particularly rapid in the muscles of animals that have died in hypoglycaemic convulsions after injection of insulin.

The physiological *rigor* caused by fatigue probably forms the basis of what Jordan and his school call viscous tonus and is an important factor in the working of most mollusc muscles. The circulation of blood in and around muscles such as the adductors of the shell or the byssal retractors of *Mytilus* is sluggish, and metabolites will probably accumulate after a short tetanus. BOZLER[80] has shown that a slight increase in the concentration of carbon dioxide very much increases the viscosity of *Helix* muscles.

Contractures— Physiological *rigor* is not the only way in which muscles can support loads without a large expenditure of energy. If stimulated by DC, *Mytilus* muscles shorten in the immediate neighbourhood of the cathode (FLETCHER[49]) and remain short for as long as the current flows. When the current is cut off, relaxation is slow, but is accelerated if the muscle is given a short stimulus by AC. WINTON[33] has investigated the dynamic modulus during this type of activity, and finds that it is intermediate between that of resting muscle and of muscle stimulated by AC, which elicits a normal twitch.

In order to investigate the effects of DC the same apparatus was used as before, with 120 v DC mains substituted for the 220 v AC mains. The same electrodes were used, making the stream of sea water from the upper electrode the cathode. Under these conditions the muscle appeared to shorten all over without any obvious local swelling like that described by Fletcher, who used contact electrodes.

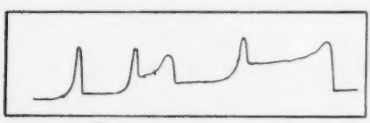


Figure 33. Effects of stimulation by DC and AC on anterior byssal retractor of Mytilus *edulis. From right to left*: **A** *DC contracture terminated by a short burst of AC stimulation,* **B** *the same, followed by a short burst of AC stimulation*

Figure 33 is a record of typical

contractions obtained by DC stimulation, and shows how a short burst of AC stimulation encourages relaxation. *Figure 34* shows the amplitude of the response to alternating stress of resting muscle, and of muscle stimulated by both DC and AC. These results agree well with Winton's.

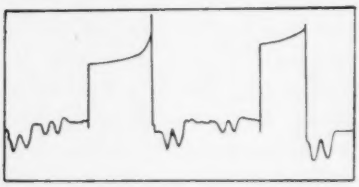


Figure 34. Effect of DC stimulation on the dynamic modulus of anterior byssal retractor muscle of Mytilus, *frequency 0·7 cycles/sec. From right to left: resting unstretched, stretch, resting after stretch, stimulation by AC, stretch, resting after stretch, stimulation by DC.*

The local contractures described by Fletcher were not accompanied by propagated electrical disturbances, but only by depolarization in the contracted region ; it seems probable that this type of contracture may always be associated with lasting depolarization of parts of the muscle membrane. The changes in mechanical properties could be accounted for if ATP were liberated in the normal way, but was then rapidly immobilized, leaving the proteins deprived of ATP in the contracted configuration. One way in which this might come about is by the persistence of the polar membrane over parts of the muscle surface and its destruction in others. Free ATP would tend to diffuse out of the active regions to become fixed at the membrane, and being thus rapidly deprived of plasticizer the contracted regions would go into a local *rigor*. Stimulation by AC would set up a propagated electrical disturbance which would break down the membrane over the whole surface, so that the proteins would be flooded with ATP all over the muscle, and relaxation could proceed normally.

Stimulation by DC is of course an artificial condition, but it seems possible that a similar effect occurs in nature in response to local depolarization at the end of the motor nerve. Contractures involving similar mechanical phenomena have been observed in the muscles of molluscs, but in the absence of any measurements of electrical potentials on the muscle surface it is difficult to distinguish such contractures from the effects of fatigue. JORDAN and POSTMA[83] have shown that electrical stimulation of the pedal ganglion within a certain range of intensities causes an increase in the viscous tonus of the foot of *Helix*, and NIEUWENHOVEN[50] has shown the same thing for the byssal retractors of *Mytilus*.

In crustacea, similar contractures have been described, and it has been demonstrated that they are not associated with any propagated electrical disturbance (MARMONT and WIERSMA [84]; KATZ and KUFFLER[85]). The contractions are slow and are confined to narrow bands near the point of entry of the motor nerves ; Katz and Kuffler

have described the appearance of local swellings. The contractures are abolished by a single propagated twitch. The muscles are also capable of normal fast twitches which are accompanied by a propagated electrical disturbance of the usual kind. The two types of contraction, fast and slow, described by PRINGLE[86] for the leg muscles of the cockroach, probably depend on a similar mechanism. The arthropods have exploited this alternative kind of contraction for the production of tonus ; a development which has made possible the evolution of muscles with a very low internal viscosity, capable of giving a rapid twitch followed by extremely rapid relaxation.

Vertebrates seem to possess a comparable system in the small motor nerves described by KUFFLER and GERARD[87]. According to KUFFLER, LAPORTE and RANSMEIER[88] the local contractures produced at the nerve endings of the small nerve system in the frog contribute to the production of tone, but there does not seem to be any reason to doubt the usual view that postural tone in vertebrates is maintained for the most part by a modified tetanus. This must set a limit to the development of very fast muscles, because any increase in the speed at which skeletal muscle can contract and relax entails a corresponding increase in the energy required for the maintenance of postural tone.

On comparing these mechanisms for the economical support of loads with what we know of the production of set in contracted tendon, it appears that their essential common feature is that the concentration of plasticizer is reduced suddenly, whereas in normal relaxation it is reduced slowly.

Latency relaxation— If a muscle under a small load is stimulated, it lengthens slightly before it contracts. At approximately constant length the tension falls by an amount corresponding to about 1/1,000 of the full isometric tension ; the whole effect occupies about 0·01 seconds (SANDOW[89] ; ABBOT and RITCHIE[90]).

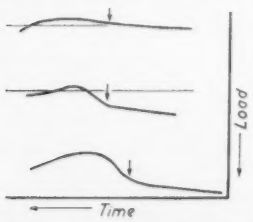


Figure 35. 'Latency relaxation' in tendons. Tracings from kymograph records of the behaviour of a 'regenerated' tendon loaded in water, to which a solution of mercury-potassium iodide is added gradually, beginning at the moment marked by the arrow. The initial load is increased from the upper to the lower tracing. All read from right to left

As has been described above (p 246), a similar relaxation can be observed in tendon, where it is apparently due to the effect of a low concentration of plasticizer in lowering the internal viscosity. A tendon or a muscle loaded with a small weight during rest will elongate very slowly. As the concentration of plasticizer increases, the rate of elongation will increase, until the concentration reaches a point at which the equilibrium load is greater than the applied load, when contraction begins. In muscle the concentration of plasticizer rises

very rapidly so that relaxation is short lived and difficult to observe, but in the model the rate of increase of plasticizer concentration can be made very slow so that the effect is easily observed (*Figure 35*). The explanation of the mechanism of latency relaxation implied in this model is in general agreement with Sandow's own suggestion that it is related to the relaxation of myosin threads on the addition of ATP (ENGELHARDT[91]). The only other explanation which has been proposed is that of SCHOEPFLE and GILSON[92], who suggest that relaxation may be due to ' a compression wave following the sudden shock which attends tension development '.

ACTION OF ADENOSINE TRIPHOSPHATE ON MUSCLE PROTEINS IN VITRO

The effects of ATP on muscle proteins *in vitro* are like those of mercury-potassium iodide on gelatin, and agree well with the kinetic hypothesis. Actomyosin gels in 0·6M KCl liquefy if ATP is added (SZENT GYÖRGYI[26]). When ATP is added to solutions of actomyosin there is an immediate decrease in viscosity, followed by a recovery to the original values as the ATP is hydrolysed (DAINTY *et alii*[93]). From measurements of the scattering of light by the solution, MOMMAERTS[94] comes to the conclusion that the fall in viscosity is due to dissociation of the actomyosin rather than to contraction of the micelles, but JORDAN and OSTER[95], using a similar technique, come to the opposite conclusion. According to MOMMAERTS[96] hydrolysis of the ATP is not an essential part of the process, as the fall in viscosity occurs just the same if the enzymatic activity of the actomyosin is poisoned with copper. Surface films of actomyosin expand on the addition of ATP (MUNCH-PETERSEN[97]).

These changes all seem to indicate a decrease in the strength of intermolecular attractions, like that which occurs in solutions of synthetic high polymers on the addition of a plasticizer ; as against this Szent Györgyi has described the syneresis of actomyosin gels in dilute solutions of electrolytes when ATP is added. This might perhaps be fitted into the picture if it is supposed that by combining with the protein, the ATP diminishes its affinity for water ; it has already been shown that the effect of mercury-potassium iodide on tendon or gelatin is not to cause extreme swelling. It is doubtful whether any large changes in water content are involved in muscular contraction *in vivo*.

ENGELHARDT, LJUBIMOVA and MEITINA[98] applied ATP to extruded myosin filaments which still retained some enzymatic activity, and showed that it made them more extensible. This has been taken as evidence that ATP is concerned with relaxation rather than with contraction (ENGELHARDT)[91], but the conclusion does not follow. The

degree of molecular orientation to be expected from extruded filaments is small, and if the only effect of ATP were to decrease intermolecular attractions as theory requires, its action on a disorientated mass would be to make it more extensible for rapid deformations and more plastic for slow ones. SHEN[99] states that the effect is abolished if the enzymatic activity of the ' myosin ' is destroyed by poisoning with silver salts, a conclusion which does not seem to agree with Mommaerts' findings on the effect of ATP on the viscosity of actomyosin solutions.

I owe a great deal to my wartime colleagues at the Royal Aircraft Establishment, Farnborough ; in particular Mr J. E. Gordon and Dr C. Gurney who have argued with me about the theory at all stages. The mathematics I owe entirely to my wife, who has worked out the integrals for me ; I must also thank Dr H. E. Daniels for his advice on the statistics. Professor A. V. Hill, Dr Gutfreund and Dr Pantin have all read the manuscript at various stages and have given me helpful advice.

REFERENCES

[1] WIEGAND, W. B. and SNYDER, I. W. *Trans. Inst. Rubber Ind.* 10 (1934) 234
[2] MEYER, K. H. *Chem. and Industr.* 57 (1938) 439
[3] WÖHLISCH, E. *Koll. Z.* 89 (1939) 239
[4] MEYER, K. H. and VAN DER WYK, A. J. A. *Experientia* 2 (1946) 1
[5] MÜLLER, F. H. *Koll. Z.* 95 (1941) 138
[6] ASTBURY, W. T. *Proc. roy. Soc. Lond. B* 134 (1947) 303
[7] MEYER, K. H. and FERRI, C. *Helv. Chim. Acta* 18 (1935) 570
[8] WOODS, H. J. *J. coll. Sci.* 1 (1946) 407
[9] — *Nature, Lond.* 157 (1946) 229
[10] LEADERMAN, H. *Industr. Engng Chem.* 35 (1943) 374
[11] ASTBURY, W. T. and WOODS, H. J. *Phil. Trans. roy. Soc. Lond. A* 232 (1933) 333
[12] WOODS, H. J. *Proc. roy. Soc. Lond. A* 166 (1938) 76
[13] HARRIS, M., MIZELL, L. R. and FOURT, L. *Industr. Engng Chem.* (Ind. Ed.) 34 (1942) 833
[14] MEYER, K. H. and HASELBACH, C. *Nature, Lond.* 164 (1949) 33
[15] ASTBURY, W. T. and BELL, F. O. *Nature, Lond.* 147 (1941) 696
[16] RUDALL, K. M. *Soc. Dy. Col. Symposium on Fibrous Proteins* Leeds, (1946) 15
[17] MERCER, E. H. *Nature, Lond.* 163 (1949) 18
[18] WÖHLISCH, E. and DU MESNIL DE ROCHEMENT, R. *Z. Biol.* 85 (1927) 406
[19] PICKEN, L. E. R. *J. Chim. Phys.* 34 (1937) 764
[20] MEYER, K. H. *Naturwiss.* 26 (1938) 199
[21] SPEAKMAN, J. B. *J. Soc. Dy. Col. Bradford* 52 (1936) 385
[22] RAMSEY, R. W. *Muscle : physics*, in *Glasser O. Medical Physics*, 1944
[23] MEYER, K. H. *Bioch. Z.* 214 (1929) 253
[24] — *ibid* 217 (1930) 433
[25] WEBER, H. H. *ibid* 217 (1930) 431
[26] SZENT GYÖRGYI, A. *Chemistry of Muscular Contraction* New York, 1947
[27] PERRY, S. V. and REED, R. *Bioch. Biophys. Acta* 1 (1947) 379
[28] RISEMAN, J. and KIRKWOOD, J. G. *J. Amer. chem. Soc.* 70 (1948) 2820
[29] WIEGAND, W. B. *Trans. Inst. Rubber Ind.* 1 (1925) 141
[30] MONGAR, J. L. and WASSERMAN, A. *Nature, Lond.* 159 (1947) 746
[31] TUCKETT, R. F. *Trans. Farad. Soc.* 38 (1942) 310
[32] ALFREY, T. and DOTY, P. *J. appl. Phys.* 16 (1945) 700
[33] WINTON, F. R. *J. Physiol.* 88 (1937) 492
[34] SCHMIDT, W. J. *Die Doppelbrechung von Karyoplasma, Zytoplasma, und Metaplasma* Berlin, 1937
[35] WHEWELL, C. S. and WOODS, H. J. *Soc. Dy. Col. Symposium on Fibrous Proteins* Leeds, 194
[36] KAYE, M. and LLOYD, D. M. *Proc. roy. Soc. Lond. B* 96 (1924) 293
[37] ASTBURY, W. T. *J. Soc. Leather Tr. Chem.* 24 (1940) 69
[38] MARCEAU, F. *Arch. Zool. Exp. Gen. S* 5 *T* 2 (1909) 296
[39] WÖHLISCH, E. *Bioch. Z.* 247 (1932) 328

[40] FAURE-FREMIET, E. and WOELFFLIN, R. *J. Chim. Phys.* 33 (1936) 55
[41] WÖHLISCH, E. *Koll. Z.* 96 (1941) 213
[42] BULL, H. B. *J. Amer. chem. Soc.* 67 (1935) 2047
[43] FISCHER, E. *Ann. N.Y. Acad. Sci.* 47 (1947) 783
[44] HALL, C. E., JAKUS, M. A. and SCHMITT, F. O. *Biol. Bull.* 40 (1946) 32
[45] HARTREE, W. and HILL, A. V. *J. Physiol.* 55 (1921) 389
[46] MEYER, K. H. and PICKEN, L. E. R. *Proc. roy. Soc. Lond. B* 124 (1937) 29
[47] RENK, F. and WÖHLISCH, E. *Pflüg. Arch. ges. Physiol.* 243 (1940) 110
[48] BUCHTAL, F., KAISER, E. and KNAPPEIS, G. G. *Acta Physiol. Scand.* 8 (1944) 16
[49] FLETCHER, C. M. *J. Physiol.* 90 (1937) 233
[50] NIEUWENHOVEN, L. M. *Thesis* Utrecht, 1947
[51] RAMSEY, R. W. and STREET, S. F. *J. Cell. Comp. Physiol.* 15 (1940) 11
[52] SICHEL, F. J. M. *Amer. J. Physiol.* 133 (1941) P 446
[53] RAMSEY, R. W. *Ann. N.Y. Acad. Sci.* 47 (1947) 675
[54] ASTBURY, W. T. *Nature, Lond.* 160 (1947) 388
[55] BUCHTAL, F. *Det. Kgl. danske Vidensk. Selskab. Biol. Meddl. B* 17 (1942)
[56] MEYERHOF, O. *Ann. N.Y. Acad. Sci.* 47 (1947) 815
[57] HØNCKE, P. *Acta Physiol. Scand.* 15 (1947) Suppl. 48
[58] FENN, W. O. *J. Physiol.* 58 (1923) 175
[59] — and MARSH, B. S. *ibid* 85 (1935) 277
[60] HILL, A. V. *Proc. roy. Soc. Lond. B* 126 (1938) 136
[61] — *ibid* 136 (1949) 195
[62] — *ibid* 136 (1949) 211
[63] — *ibid* 136 (1949) 220
[64] — *ibid* 136 (1949) 242
[65] HILL, A. V. *Bioch. Biophys. Acta.* 4 (1950) 4
[66] GASSER, H. S. and HILL, A. V. *Proc. roy. Soc. Lond. B* 96 (1924) 398
[67] SEEMANN, J. *Pflüg. Arch. ges. Physiol.* 106 (1905) 420
[68] HILL, A. V. *Proc. roy. Soc. Lond. B* 100 (1926) 108
[69] LEVIN, A. and WYMAN, J. *ibid* 101 (1927) 218
[70] HUMPHREY, G. F. *Physiol. comp. et Oecol.* 1 (1948) 89
[71] KATZ, B. *J. Physiol.* 96 (1939) 45
[72] HILL, A. V. *Proc. roy. Soc. Lond. B* 135 (1948) 446
[73] PETIT, J. L. *Arch. Int. Physiol.* 34 (1931) 113
[74] HOGBEN, L. T. and PINHEY, K. F. *J. Exp. Biol.* 4 (1926) 196
[75] CASPERSSON, T. and THORELL, B. *Acta Physiol. Scand.* 4 (1942) 97
[76] ENGSTRÖM, A. *ibid* 8 (1944) 137
[77] BRONK, D. W. *J. Physiol.* 69 (1930) 306
[78] FENG, T. P. *Proc. roy. Soc. Lond. B* 108 (1931) 522
[79] BRONK, D. W. *J. Cell. Comp. Physiol.* 14 (1932) 429
[80] BOZLER, E. *Z. vergl. Physiol.* 14 (1931) 429
[81] HILL, A. V. *Proc. roy. Soc. Lond. B* 109 (1931) 267
[82] BATE SMITH, E. C. and BENDALL, J. R. *J. Physiol.* 106 (1947) 177
[83] JORDAN, H. J. and POSTMA, N. *Proc. nederl. Akad. Wetensch., Amsterdam* 44 (1941) 1196
[84] MARMONT, G. and WIERSMA, C. A. G. *J. Physiol.* 93 (1938) 173
[85] KATZ, B. and KUFFLER, S. W. *Proc. roy. Soc. Lond. B* 133 (1946) 374
[86] PRINGLE, J. W. S. *J. Exp. Biol.* 16 (1939) 220
[87] KUFFLER, S. W. and GERARD, R. W. *J. Neurophysiol.* 10 (1947) 383
[88] — LAPORTE, Y. and RANSMEIER, R. E. *ibid* 10 (1947) 395
[89] SANDOW, A. *J. Cell. Comp. Physiol.* 24 (1944) 221
— *Ann. N.Y. Acad. Sci.* 47 (1947) 895
[90] ABBOT, B. C. and RITCHIE, J. M. *J. Physiol.* 107 (1947) 5
[91] ENGELHARDT, W. A. *Advances in Contemporary Biology*, 2nd Ed. 1941
[92] SCHOEPFLE, G. M. and GILSON, A. S. *J. Cell. Comp. Physiol.* 27 (1946) 105
[93] DAINTY, M., KLEINZELLER, A., LAWRENCE, A. S. C., MIALL, M., NEEDHAM, J., NEEDHAM, D. M. and SHEN, S. *J. gen. Physiol.* 27 (1943) 355
[94] MOMMAERTS, W. F. H. M. *Ark. Mineral. och Geologi* 19 *A* (1945)
[95] JORDAN, W. K. and OSTER, G. *Science* 108 (1948) 188
[96] MOMMAERTS, W. F. H. M. *J. gen. Physiol.* 31 (1948) 361
[97] MUNCH-PETERSON, A. *Nature, Lond.* 162 (1948) 537
[98] ENGELHARDT, W. A., LJUBIMOVA, M. N. and MEITINA, R. A. *Compt. Rend. Acad. Sci. U.R.S.S.* 30 (1941) 644
[99] SHEN, S. *Ann. N.Y. Acad. Sci.* 47 (1947) 875